FIELDS OF PSYCHOLOGY

AN EXPERIMENTAL APPROACH

Edited by

ROBERT H. SEASHORE
Northwestern University

With Contributions from

CLAUDE E. BUXTON
DONALD B. LINDSLEY
HARRY F. HARLOW
BETH L. WELLMAN
DAEL L. WOLFLE
E. G. WILLIAMSON
HARRISON MUSGRAVE
HAROLD G. SEASHORE
PAUL R. FARNSWORTH
EDMUND S. CONKLIN

NEW YORK
HENRY HOLT AND COMPANY

July, 1949

PRINTED IN THE
UNITED STATES OF AMERICA

Preface

THE title *Fields of Psychology: An Experimental Approach* indicates that experimental methods have now been extended to cover all of the special fields of psychology. As evidence we shall present a series of descriptions of modern research which will include all gradations of methods from the classical single variable type of problem to the latest multiple variable types such as are used in factorial analyses. We shall also introduce the reader to the psychologists' various places of work, from the sound-proof, electrically shielded darkroom laboratories to the schools, clinics, industries, studios and playing fields which are often more representative situations for the study of behavior in everyday life.

Our collaborative plan reflects the remarkable degree of specialization in psychology, so great that even other psychologists need a series of experts to keep them in touch with the rapid progress of research in the various fields. Having selected the collaborators for their active contributions to each field, the editor has frequently invited each to present his own work as a part of the series of experiments representing that field of research. Individual variations in procedure such as the description of one or several experiments in a single chapter simply reflect the fact that research areas are sometimes cultivated intensively by a single individual while in other cases our knowledge of an area comes only from the successive contributions of several different investigators.

Aside from the editor, each contributor is responsible only for the points of view expressed in his own section. This, however, has represented no great problem since the experimental approach usually provides empirical tests for the resolution of differences in theory.

This volume has been designed to promote a better understanding of the progress of modern psychology by surveying representative experiments in each special field. If each of us has enabled others to appreciate the progress in his chosen field we shall have achieved a major aim. It would be too much to expect each reader to find an equal degree of interest in every special field, but at least we believe that this affords a sample by which to test one's interest in each field.

We wish also to acknowledge the courtesies of the various authors and publishers who have graciously permitted us to use the materials from original researches as acknowledged in each section of the book.

R. H. S., EDITOR.

Northwestern University
Evanston, Illinois

Contents

PART ONE. INTRODUCTION

By

Robert H. Seashore

Northwestern University

PART TWO. GENERAL EXPERIMENTAL PSYCHOLOGY

By

Claude E. Buxton

Northwestern University

PART THREE. PHYSIOLOGICAL PSYCHOLOGY

By

Donald B. Lindsley

Brown University

PART FOUR. COMPARATIVE PSYCHOLOGY

By

Harry F. Harlow
University of Wisconsin

PART FIVE. HUMAN DEVELOPMENT

By

Beth L. Wellman
State University of Iowa

PART SIX. EDUCATIONAL PSYCHOLOGY

By

Dael L. Wolfle
University of Chicago

PART SEVEN. VOCATIONAL GUIDANCE

By

E. G. Williamson

University of Minnesota

PART EIGHT. INDUSTRIAL PSYCHOLOGY

By

Harrison Musgrave

Southern Counties Gas Co. of California, Los Angeles

PART NINE. AVOCATIONAL PSYCHOLOGY

By

Harold G. Seashore

Springfield College

PART TEN. SOCIAL PSYCHOLOGY

By

Paul R. Farnsworth

Stanford University

PART ELEVEN. ABNORMAL AND CLINICAL PSYCHOLOGY

By

Edmund S. Conklin
Indiana University

PART TWELVE. SYSTEMATIC PSYCHOLOGY

By

Robert H. Seashore

Part One

INTRODUCTION

By

ROBERT H. SEASHORE

Northwestern University

CHAPTER I

An Experimental Approach to the Fields of Psychology

THE PLAN of this book arose from the experience of the editor in visiting various professional colleagues at other institutions and being conducted on trips through their laboratories, clinics and personnel offices. The great profit derived from such opportunities to see what was going on in leading institutions of the country and how it was being done suggested an attempt to extend this experience to students by securing *a still more representative sample of current psychological investigations in each of the special fields* and by having an expert in each field interpret the methods and findings in print.

The present volume in no way displaces the usual type of general introductory text book in psychology, being designed instead to supplement their more theoretical approach and their briefer mention of research findings by supplying fairly complete summaries of representative experiments in the various specialized fields. Neither do we attempt to compete with laboratory manuals which are designed to give concrete experience with a few typical experiments in the general field. Instead we hope to enable a student to see how these same methods which he has used in the laboratory or class demonstrations have also been applied to the special fields which he cannot find time to cover by himself. We are here attempting to show in greater detail how psychologists actually go about their experiments, so as to make possible a much fuller understanding of the results and their interpretation.

In contrast with other volumes entitled *Fields of Psychology,* we have not attempted to list all of the principal subtopics or to give a bird's-eye view of the whole field, valuable

as these aims may be. Instead we have attempted to show in considerable detail just how some of our major investigations have developed, so that the student can judge for himself as to their adequacy, and their degree of interest and value for him in choosing further and more specialized courses. If our readers are given a better picture of progress in some of the more important lines of research in modern psychology and of the interrelation of the various fields, we shall have achieved our major aims.

In the early stages of experimental psychology a single investigator could write a fairly comprehensive general text or experimental manual from his own research in several areas and from reading or hearing firsthand reports in all of the other areas. Thus in the 1890's one of the writer's friends could fairly say that he knew personally practically all of the leading psychologists of Europe and America and had read almost every important book which appeared in modern psychology. Within the next few decades psychology grew so rapidly and spread out in so many different directions that the pioneers who explored each new area lost touch with one another, each seeking the advantages of specialization through division of labor. Eventually the "pigeonholing" of scientific fields went so far and the interrelations between fields were so little understood that each group tended to see everything in terms of its own background and attempted to make that background *the* basic interest and approach of all psychology. But as these specialized areas developed further, they again made contact with one another and many psychologists began to see how their separate contributions could be fitted together in a single eclectic [1] science of psychology.

In the present stage of cooperative development the older schools of thought are seen to represent primarily a difference of interests in various problems, all of which are necessary for a complete psychology.

[1] The term eclectic in science refers to a viewpoint which attempts to combine the best contributions from all schools of thought.

The various specialists now find it profitable to cooperate with one another in research, pooling their information, methods, and tests in the study of their common problems. Practically all large research organizations work in this way. One of the marks of a maturing science is the more frequent appearance of multi-authored articles and books, a trend which is very noticeable in present-day psychology. In this way specialists not only advance the development of their own older fields, but create or discover new "bridge sciences" such as the new "avocational psychology" between the recreational fields of esthetics and athletics on one side and psychology on the other. As a psychologist once said after attending a series of scientific meetings, many of our research interests and lines of progress are now being concentrated in these neglected "cracks" between the older fields of science.

The Classification of the Fields of Psychology

In order to understand the widespread developments of modern psychology we have chosen to summarize a series of experiments from each of its special fields. Although psychology may be classified in various ways, the following list of fields is reasonably complete and in accord with present usage.

1. GENERAL EXPERIMENTAL PSYCHOLOGY

This is a study of our sensory, affective, motor, and intellectual activities in their simplest and most basic forms, as usually described in elementary general texts and experimental manuals. Experimenters in psychology seek to answer questions about our experience and behavior such as those listed below. The illustrations of typical problems are taken from the section on general experimental psychology.

A. Qualitative.

1. What *kinds* of activities occur? Illustrations: the experience of apparent movement as described in Ch. 2, or the kinds of bodily activities in the emotion of startle, as in Ch. 4.

2. In what typical *patterns* do activities occur? Illustrations:
    a. at a given time: the frame of reference for apparent motion (Ch. 2) and the "startle" pattern of activities, face, arms, trunk (Ch. 4.).
    b. during their characteristic rise and fall over a longer period of time: the development of a conditional response (Ch. 6), learning to read (Ch. 21), development of intellectual, motor and personality characteristics (Chs. 15 to 19).

B. Quantitative.
1. How much, or what degree, of each kind of activity appears, and how is it related to other activities? Illustration: the interrelations of mechanical abilities (Ch. 5).
2. What is the numerical relationship between varying amounts of stimuli and responses in a given type of organism? Illustration: the minimal stimuli for hearing (Ch. 3).

Both qualitative and quantitative methods are necessary to the complete study of anything. If we study an activity at a single stage of development, such a study is called a *cross-sectional* or *static* analysis. This is usually necessary as a first step in any science. If we later trace the development through all the characteristic steps, this is called a *longitudinal* or *genetic* study. Our experiments in Ch. 6 on human conditioning and in Section V on developmental psychology are of this sort. Both cross-sectional and genetic methods are necessary to the complete study of anything.

In psychology the study of *experience* (implicit or covert activity [2]) happened to grow up first and quite independently of the study of explicit or *overt behavior*. There are many lines of evidence to show that experience and behavior are simply two degrees of the same general kind of activity, and both are included in the fields of psychology.

[2] Implicit activity is that which cannot be observed by other people without instruments, usually because the activity is internal or small in size. Thus the word *experience*, as used here, has a more precise meaning than it has in common speech. Experience has a subjective connotation.

Qualitative study of experience (sensory, affective, and intellectual processes) is called *introspection* but this procedure might just as well have been called *inspection* as in the other sciences. Introspection is simply observation and description of any experience which may be produced under controlled conditions, and may be verified through the study of similar phenomena by other self-observers to see how much is common to all persons and how much varies with individuals or from one time to the next in the same person.

In qualitative studies of overt behavior the processes of observation and description are essentially the same as in the study of experience, except that overt activities are somewhat easier to record, e.g., by photography, as in Ch. 4 on the "startle" responses.

Quantitative studies of either experience or behavior are called *psychometrics,* or the measurement of psychological processes.

Psychometric studies of experiences are historically known as *psychophysics*, or the observation (measurement) of the numerical relations between certain aspects of psychological activities and the physical stimuli which set them off. There is as yet no corresponding specialized term for the measurement of overt behavior, but the term *motometry* (from motor measurements) might be coined as an appropriate parallel.

Genetic methods may study developments within the lifetime of a single individual, in which case they are called *ontogenetic* methods. If the study covers the development of an evolutionary series of individuals, groups or species of animals or people, it is called *phylogenetic.* The tracing of the development of behavior in family lines is in the field of *heredity.* The development of behavior among racial, national or other groups is a part of *social psychology* and anthropology, while the study of the evolutionary development of behavior from earlier to later animal species is in the field of *comparative* or *animal* psychology. Animal behavior may also of course be studied in a single animal or species purely for its own interest, by cross-sectional methods.

Genetic studies may be either *forward tracing,* i.e., starting at the beginning and tracing each step as it develops, or they may be *backward tracing,* i.e., attempting to infer from the nature of present behavior what situations must have occurred in order to bring it about. Examples of forward-tracing studies of individual behavior deal with maturation, or the development of behavior primarily through anatomical growth, as compared to learning (including forgetting), the modification of behavior by experience. Examples of backward-tracing studies are the clinical methods of diagnosing the nature and origin of a patient's difficulties after they have fully developed.

By way of summary, a rough description of *general psychology* would be: the systematic study of sensory, affective, intellectual and motor processes, including the identification of their contents and their functions, the characteristic patterns in which they occur, both qualitative and quantitative, and both their cross-sectional and developmental aspects.

The terms "pure" and "applied" psychology represent only two different emphases: (1) the search for *basic phenomena and methods,* as compared to (2) the *practical use* which is made of these findings in various aspects of everyday life. So-called "pure" and "applied" psychologists contribute so much to each other's interests that there is no point in trying to distinguish them further. Any field of psychology may be studied with an emphasis on either its pure or applied aspects, or on both.

2. PHYSIOLOGICAL PSYCHOLOGY

When such processes as seeing, feeling, thinking, or acting are analyzed in terms of the physico-chemical action in the sensory, neural, glandular or muscular organs involved, the study is called *physiological psychology.* Originally this term included all of experimental psychology because the earliest workers came from the field of physiology, extending their related problems and methods to psychology. It is sometimes said that psychology describes the activities of the organism as a whole, while physiology studies the action of its compo-

nent organs. The roughness of this distinction as to respective emphases brings out the fact that these fields of knowledge are not sharply separated, but rather represent certain patterns of interests marked off by scientists who have been forced to specialize in order to be efficient.

3. GENETIC PSYCHOLOGY

Whereas general psychology has traditionally emphasized the cross-sectional studies of human behavior or at most the development of behavior over a short period of time, genetic psychology includes studies of development, in either human or lower animals, over any period of time, from the whole course of evolution to a short period of individual learning.

That branch of genetic psychology which involves the broader comparison of several species, e.g., rats, apes, men, is called *comparative psychology.*

a. *Comparative psychology* may be studied for the knowledge of behavior of the lower animals themselves or as a simple approach to the study of more complex human behavior. The shorter life cycles of many lower animals permit more rapid developmental studies than in man, which are useful in the studies of inheritance, for example, while the possibilities of drastically controlling their environments or operating on their bodily structures permit many studies which would be either difficult, dangerous, or otherwise undesirable for human subjects.

b. *Human genetic psychology.* The stages in human development are usually classified into the foetal or pre-natal, infancy, pre-school, grade school, adolescent, adult, later maturity, and the senescent. Child psychology is a grouping of the periods from infancy to adolescence while most general psychology is concerned with the mature behavior of adults.

4. EDUCATIONAL PSYCHOLOGY

Another name for this field might be that of applied human genetic psychology, and as a matter of fact the interests of the genetic and educational fields overlap greatly. The applied

genetic field emphasizes the ***most effective means for bringing about various socially desirable developments*** in human behavior. As such it begins with clinically determined needs of the child and comes in contact with all of the social sciences which study the environment in which the child is to live and work. The major mode of development is learning.

At present our educational systems emphasize the development of general skills, and those required by a few professions. More specific education in various trades and less highly skilled vocations has been left largely to employers in each field, but in certain areas this type of service is now being taken over by the public or private vocational schools. Our educational system has also until recently neglected avocational skills such as art, music, and athletic recreations for the general student. Educational psychology also strives to discover individual differences among people and to adapt its programs to the individual.

5. DIFFERENTIAL PSYCHOLOGY (INCLUDING PERSONNEL)

The laws or general principles of psychology are really the averages or central tendencies derived from the study of a wide range of individual variations. General psychology tells us what the average individual is most likely to do in given situations, while differential psychology goes further in attempting to predict how a given individual will deviate from the average of others. One applied aspect of differential psychology is the study of personnel, which is in turn a major interest of educational psychology in ***vocational guidance*** of individual students and of industrial psychology in ***vocational selection*** of the most promising applicants for employment.

Differential psychology is also interested in the organization or interrelations of human abilities, so that we can know what groups of abilities may be studied to predict success in various specialized fields. In this volume we shall study differential psychology as applied to specific problems in the other fields.

6. VOCATIONAL PSYCHOLOGY (BUSINESS AND INDUSTRIAL)

Industrial psychologists can now be of service to businessmen at almost every step in their work. Beginning with the plans for producing goods, the expert in consumer research can survey the needs and preferences of prospective customers so as to aid in designing a widely desired and easily marketable product. With the aid of industrial engineers, the industrial psychologist can assist in designing a plant and its equipment in accordance with labor-saving principles derived from time and motion studies. The personnel specialist can aid in the selection and training of employees, and advise the management in relation to the motivational values of various programs for wages, advancement, supervision, safety, social security, etc. Finally advertising specialists can assist in evaluating advertising and selling campaigns by giving them an experimental tryout on a small scale and correcting minor faults before the investment of large sums of money on a major sales campaign. Consumers could likewise be assisted in working out methods of purchasing which would be more effective in the satisfaction of their needs, but this field is still in its infancy.

7. AVOCATIONAL (RECREATIONAL) PSYCHOLOGY

The same methods of selecting individuals and training them in effective work methods which have been found useful in educational and vocational psychology can also be employed in selecting and developing our avocations, such as the appreciation of the arts, music, literature, and other esthetic fields; recreation through athletic and other games; and hobbies such as gardening, carpentry, and handcrafts, all of which may also be the vocations of other people.

8. SOCIAL PSYCHOLOGY

Whenever other people are a part of our environment, and whenever we join with other people in any activity, we are touching upon social psychology. Compared with our evolution along biological lines, our development along social lines is still

in its infancy. The social sciences are less highly developed than the natural sciences, and social psychology is no exception. The recent depression and the second world war have forcibly impressed upon us the grave importance of our social problems, but we have as yet relatively few significant experiments in this field. Until recently, social psychologists have hesitated to attack the complex problems of everyday life. But under the New Deal, all sorts of gigantic social experiments are being tried out, often very hurriedly and with unforeseen consequences. Only in the last few years have we begun to utilize the principle of the "pilot experiment" in which a profound change is tested on a small scale as in the local testing of various federal surplus food stamp plans before committing the whole nation to the possible advantages and dangers of such new plans.

A great many psychologists believe that our greatest opportunities for progress in psychology during the next few decades lie in the neglected field of social psychology.

9. CLINICAL AND ABNORMAL PSYCHOLOGY

Technically, any deviation from average or normal behavior is *abnormal,* and is a part of the fields of clinical and abnormal psychology. We include not only disorders involving emotional and rational unbalance, such as the neuroses and psychoses, but also extreme deviations in intelligence, such as feeble-mindedness or genius.

Abnormal psychology was first studied by the clinical method, in which one attempts to work backward from a person's present difficulty to its origin in some earlier disturbing situation. From this it is often possible to work out a plan for his readjustment. Insofar as clinical methods involve careful and systematic analyses, they are inductive and approach the accuracy of the usual forward-tracing types of experiment. Recently we have been able to test some of our clinically derived theories as to the origin of abnormal behavior by setting up controlled experimental situations to produce or eliminate abnormal behavior in animals and to a lesser degree in human

beings. Perhaps the most significant aspect of abnormal psychology for most of us is the insight which it affords into our own minor behavior difficulties; for the serious deviations which we find in the psychoses and neuroses are often merely exaggerations of the minor idiosyncrasies which appear in the average person.

10. SYSTEMATIC PSYCHOLOGY

Having set up a division of labor for research in the specialized areas of psychology, we soon felt the need for some means of integrating these diverse interests and findings. A system of psychology attempts to organize the various aspects of the science in some logical and convenient order comparable to that achieved by a catalogue or filing system which makes information readily available. A system is also an index of working hypotheses for answers to problems not yet solved, suggesting what to look for and what methods of investigation are likely to be most effective in each case. Various "schools of thought" or viewpoints such as structuralism, functionalism, behaviorism, psychoanalysis, are really fragmentary systems covering only small parts of the whole area of psychology, and all of them should ultimately be fitted together to explain the field as a whole. Systematic psychology should be the master chart of progress for the unification of special fields and viewpoints in psychology.

Part Two

GENERAL EXPERIMENTAL PSYCHOLOGY

By

CLAUDE E. BUXTON
Northwestern University

CHAPTER 2

The Frame of Reference[1]

THE NOTION of the frame of reference is a somewhat novel way of indicating what psychologists have known or assumed for many years: that responses to any situation are governed by not one or two but many stimuli. Ordinarily, when we perceive or learn or make a judgment, we *do* react to a few outstanding stimuli, but this fact by no means rules out the effects of other stimuli present at the time. Nor does it rule out memory for what has happened before in a given situation. We usually stop our car at a stop sign even though no car is approaching or no policeman can be seen. It is quite likely that we are then responding not only to the sign itself and the command it carries, but to our realization that the principle on which the sign is based is good, or to our memory for the fact that last week we did not stop and were thereupon arrested by a policeman whom we had not noticed.

The frame of reference may be defined as the background of stimulation which influences our behavior in a particular situation. It may include external or internal stimuli other than the outstanding ones. It may include ideas or memories. But an important assumption is implicit in our simple definition, namely, that the effects of any given stimulus upon a person are ***not independent*** of the effects of other stimuli. The concept of the frame of reference is intended to add to our understanding of why a person does a certain thing by calling attention to the background reasons for it. And we assume that the thing he does (stopping at the stop sign, in the example above) really ***is***

[1] Much of this chapter is based upon M. Sherif's *The Psychology of Social Norms,* Harper's, New York, 1936, xii + 209 pp., and upon the same author's "A Study of Some Social Factors in Perception," *Arch. Psychol.,* 1935, No. 187, 60 pp.

influenced by or is dependent upon background stimuli (memory of being arrested previously, etc.).

Some examples of how the frame of reference operates. Observation and experiment have given us many examples of ways in which the frame of reference governs behavior. If we present two hungry men with a freshly broiled pork chop we may at first be at a loss to understand why one turns away from it with loathing and the other eats it with gusto. Upon questioning the two men we may find that one is a Mohammedan, who has been taught that pork is filthy and anyone who eats it is debasing himself; the other man may be a Christian, who has had no such training. Or we may find that the first man has at some previous time over-eaten on pork chops so that even when hungry he has no liking for them. In either of these two cases the seemingly paradoxical fact that one man eats and one turns away may be understood quite fully if we realize that the past experience or training of the two is quite different. The presence of the pork chop may be the outstanding stimulus, but the response to it is governed also by memory for what has happened previously, and this we refer to as background stimulation—the frame of reference.

In psychophysical experiments we study discrimination between stimuli which are parts of the same quantitative series. For example, one may judge lifted weights which vary in a series from 80 to 120 grams by four-gram steps, simply on the basis of how they "heft" when one is blindfolded. Such a judgment is easier to make if first a standard weight is presented. The subject can then estimate with some accuracy whether the comparison weight is lighter or heavier, and by how much. The interesting fact is, however, that after a few rounds of judgments it is not necessary to present the standard stimulus before every comparison stimulus. The "relative" judgment can be changed to an "absolute" one. This is because a *subjective* scale is established by the person making the judgments, and the comparison stimuli can be judged with reference to this subjective scale, rather than with reference to the standard weight.

Under these conditions, when the subject has practiced enough, a certain weight (e.g., 96 grams) may seem *heavy* within a certain series (e.g., 84-100 grams), for which he has built up a subjective scale. This same weight will gradually become a relatively *light* one if the whole series is shifted from 84-100 grams to 92-108 grams. The same physical object seems heavy, then light, depending upon what kind of background (total series) is set up. A different example of the same fact is afforded by college basketball coaches who require their players to wear flapping galoshes during class hours. The theory seems to be that when they put on basketball shoes these will seem very light and easily controlled, thereby enhancing the footwork of the players on the floor. The difference in weight between ordinary street shoes and basketball shoes is not great enough to produce the subjective impression of lightness when the latter are donned. But if a subjective impression of unwieldiness is built up while galoshes are worn, the basketball shoes seem very light within the weight "series" of footwear.

We are accustomed to see our college friends and acquaintances take part in campus theatrical productions. We do not think it peculiar, when we see them on the stage, if their diction, or bearing, or other characteristics change markedly, because this is necessary for "good theater." But what is acceptable in this setting may not be acceptable outside, and thus it is that we object to the behavior of a person who is "stagey" in the sorority house or class room. The different judgments which we render on the same kind of behavior in these two different situations are based on our knowledge or expectation of what ordinarily goes on in the theater as contrasted with everyday behavior outside it. The knowledge or expectation in this case may be considered part of the frame of reference within which we respond to or judge stage-like behavior.

In some colleges and universities in the country the honor system is used for all examinations. Under such a system it may be considered bad form for a student to crib or otherwise take advantage of his fellows. In the cases where this system

is most successful, the students have built up a tradition of honesty in examinations, and they themselves take the major share in enforcement of the tradition. But on many other campuses where the honor system has not developed, instructors stay in the room while examinations are being written, and on occasion act as policemen to prevent cheating. In this case, students sometimes do not consider cheating as being very bad. The major difference between a campus where this is true and a campus where the honor system works effectively is likely to be in the attitudes and the standards (frame of reference with respect to examination conduct) which are handed down to each generation of students.

It can now be seen that we must use our first definition of the frame of reference in a fairly inclusive way. It must take account of *all* the background stimuli which influence responses to any outstanding stimulus. Not only may there be *external* background stimuli—lights, sounds, objects, words—in addition to, or surrounding, the major stimuli; there may also be internal stimulation from ideas, images, and the sorts of complex ideas that we call standards or attitudes or expectations. There are thus two major parts of the frame of reference. One comes from past experience or training, by way of memory; one occurs in the present, by way of sense-organ stimulation (external or internal) in addition to what seem to be the outstanding stimuli.

How does the frame of reference develop? One of the first questions concerning the frame of reference is: where does it come from? Insofar as the framework is dependent upon memory, the answer is fairly straightforward. We constantly learn new things, from our own experience, or from those around us. Within a given social setting, certain ideas or ways of thinking, certain standards of behavior, are taught or learned in unceasing succession. The term frame of reference simply symbolizes how these ideas and standards come to influence behavior. Within our own culture, the way the two sexes are supposed to behave among themselves or toward each other is taught from early youth. It is no cause for wonder, then, that

by the time people are old enough to go to college, they have fairly definite ideas of what is acceptable and unacceptable behavior, either for themselves or for others. Nor is it surprising that such notions should serve as a background which governs judgments about how to behave on a "date" or at a smoker or at a "prom." The tendency to label others as radical or conservative in political and social beliefs can likewise be interpreted in part in terms of the framework of ideas which any given student gets from a firm Republican father or a good Socialist home.

The rôle of background sense-organ stimulation in a frame of reference is more difficult to describe. Controversial physiological considerations must be put aside until we are better able to describe what the background stimuli are in a given situation. We can, however, begin to determine what influence each has upon the behavioral outcome. Later in this chapter we shall present one research which attempts to do this.

How obvious is the frame of reference? In some cases, as in the judgment of weights in various series, the frame of reference is fairly obvious. In other cases, especially in judgments of ethical or moral situations, the frame of reference is more complex and not so clear-cut.

It need not be assumed, for example, that a person is fully conscious of why he reacts violently against someone who mistreats a dog. The experience which is responsible for part of his frame of reference in this case may lie far back in childhood, where someone mistreated one of his pets—the incident itself may be forgotten, and only a vague set of "liking all animals" or "disliking anyone who is not kind to animals" may remain to guide his present reaction. Nor need a relatively sophisticated and urbane individual be fully aware of why he does not like slighting and over-critical remarks about religion. He himself may say that he is no longer interested in such things. Nevertheless, some of the experiences of earlier days may exist in memory in some vague form which makes these remarks seem unpleasant rather than amusing now. In either

of these two examples there may be dozens of past occurrences rather than single incidents which are effective.

How necessary is the frame of reference? Since our present behavior depends upon (is modified in terms of) our past experience, and since present stimulation usually is complex, it probably is safe to say that some frame of reference is effective for every act. The framework may be extremely significant or relatively unimportant; it may be conscious or unconscious; it may be simple or complex.

If this be granted, there still is the question as to how necessary such a framework really is. We shall use the experimental method in answering the question. It is possible to create situations in which the response to certain controlled stimulation is not influenced by past experience *or* other stimulation during the experiment (or at least is relatively uninfluenced). We may observe just what a subject in such a situation will do—whether he continues to get along without an effective frame of reference, or whether he constructs or acquires one which is suitable and then uses it.

Such an experiment has been carried out by Sherif. He made use of what has been called the autokinetic phenomenon. This phenomenon can be demonstrated in the following way: a person is seated in a room which is completely dark except for a pin-point of light some 15 feet in front of him. If he fixates the spot of light steadily for a short while (say 30 seconds) it will *seem* to begin to move in one direction or another. One might call this event an illusion. At any rate, the visual *stimulus* actually does not move, although the subject may not realize this. But there definitely is apparent movement in a continuous path, more frequently up and to the right than in any other direction. A subject may be given many trials in this situation, or several subjects may be used together, and the movement will still seem to occur for each subject (see Fig. 1).

We are fairly sure that the frame of reference here, at least in the beginning, is lacking or is relatively ineffective for several reasons: (1) Few, if any, subjects have ever heard of the phenomenon—thereby eliminating past experience in such situ-

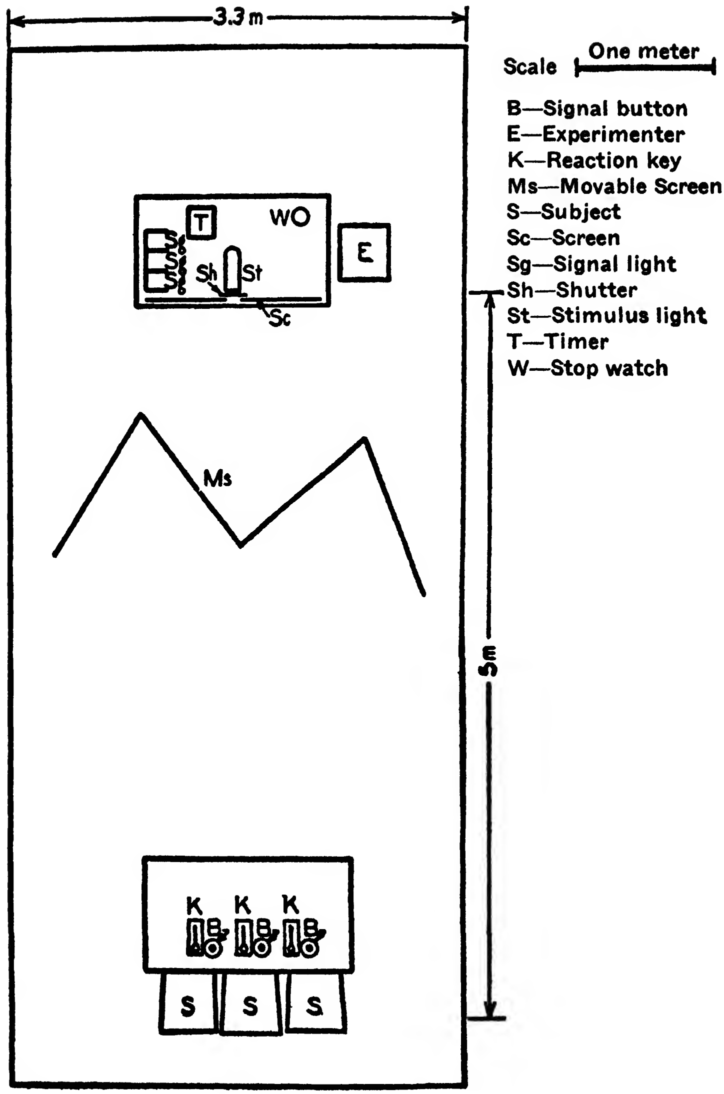

3.3 m
Scale
One meter
B—Signal button
E—Experimenter
K—Reaction key
Ms—Movable Screen
S—Subject
Sc—Screen
Sg—Signal light
Sh—Shutter
St—Stimulus light
T—Timer
W—Stop watch
T
WO
Sg
Sg
Sg
Sh
St
Sc
E
Ms
5m
K
K
K
B
B
B
S
S
S

FIG. I.

ations. (2) There is no other visual stimulation which might enter and influence judgments of how far the pin-point has seemed to move. (3) So long as the experimenter does not tell the subject anything about the illusion, incidental auditory stimuli have little or no effect upon its extent. (4) There is some reason to believe that proprioceptive stimulation (from eye-movements) does not affect the illusion greatly, but even if it does, it is in a relatively unconscious way.

The amount of movement which the subject reports might be described as altogether subjectively determined, in that the physical stimuli in the situation are constant, and the extent of the perceived movement (the illusion) depends upon the individual judgment. Actually the physical situation is ambiguous —it does not dictate what the subject shall report. Our experimental problem has thus been rephrased slightly. It becomes one of the extent to which perception of movement is modified or controlled by a frame of reference developed during the course of the experiment. We may also determine just what *kind* of framework does develop.

A single subject was brought into the room, and seated in a chair at a table. About five meters away, behind a screen, was the stimulus light. The following instructions were given:

> When the room is completely dark, I shall give you the signal ready, and then show you a point of light. After a short time the light will start to move. As soon as you see it move, press the key. A few seconds later the light will disappear. Then tell me the distance it moved. Try to make your estimates as accurate as possible.[2]

The subject's head was placed in a holder, to minimize head movements, the room lights were turned out, the screen removed, and the ready signal given. A photographic shutter (see Fig. 2) was then opened, exposing the pin-point of light. When the subject pressed his reaction key (K, Fig. 1), signifying that apparent movement had begun, the timing device was automatically set in motion. At the end of two seconds in all ex-

[2] M. Sherif, "A Study of Some Social Factors in Perception," *Arch. Psychol.*, 1935, No. 187, p. 23.

periments, the timer produced a soft click, at which time the experimenter closed the shutter so that the light disappeared.

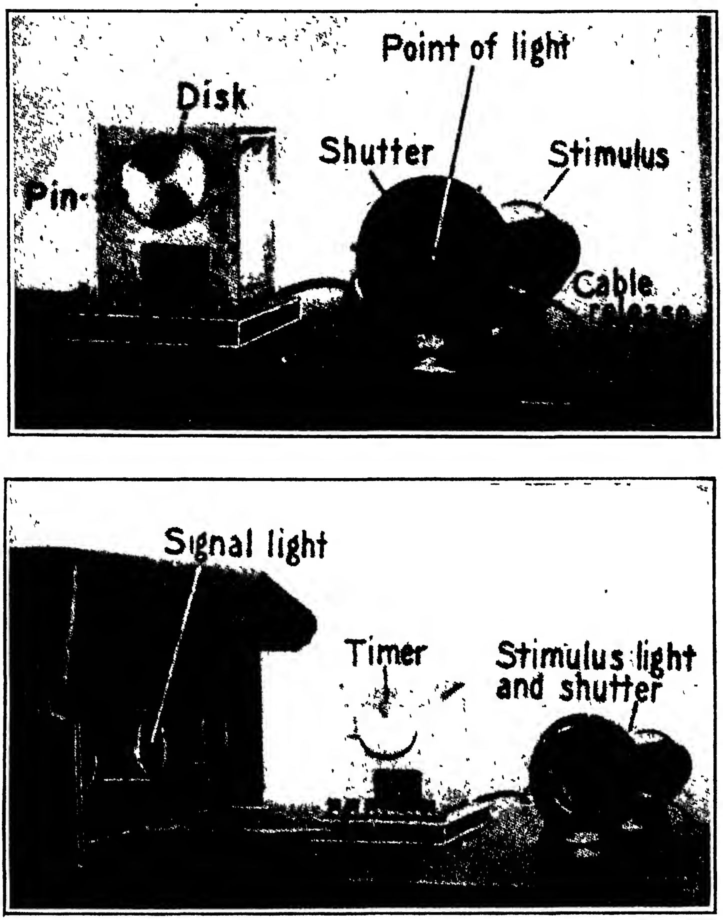


FIG. 2. (*Above*). Apparatus for individual trial with screen removed.
(*Below*). Apparatus for group experiments with screen removed.

The subject then made his report. If no movement was signaled by the pressing of the reaction key, for a period of 30 seconds after the light had been exposed, the shutter was closed and the distance of movement recorded as zero for that trial.

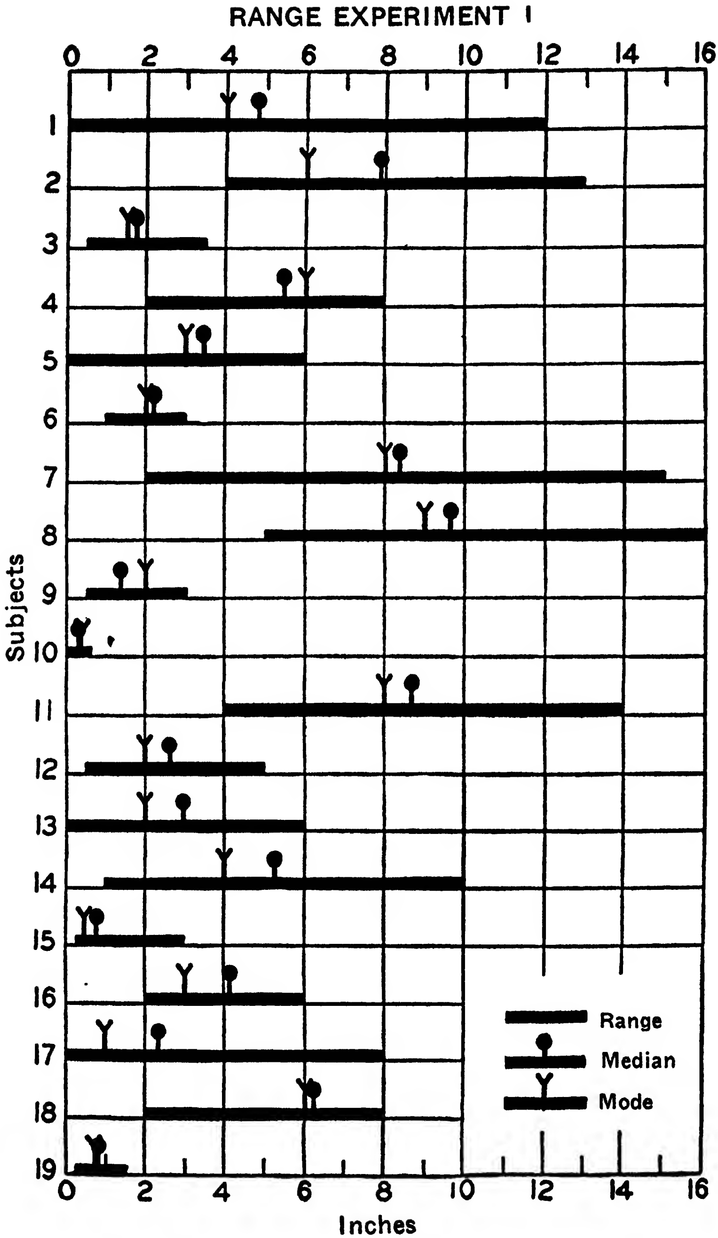

RANGE EXPERIMENT I
0 2 4 6 8 10 12 14 16
Subjects
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19
Range
Median
Mode
Inches

FIG. 3.

This very rarely happened; in fact, the key usually was pressed within five seconds after the exposure of the light had begun.

One hundred judgments were obtained from each subject in the first experiment. The results are shown in graphical form in Fig. 3. The quantitative data there shown tell us: (1) that during the course of 100 trials, each subject established for himself a range of distances through which the light might seem to move. Not only was the range of extents of movement fairly characteristic of each subject, but (2) the median and modal reports of amount of movement fall relatively close together in most cases. This is an indication that each subject developed his own standard or typical extent of movement. Though it is not shown here, the extents in successive trials ranged about this standard in a chance manner, with the total distribution of extents for the 100 trials forming an approximately bell-shaped curve.[3]

At the end of the experiment each subject was asked the following questions: [4]

1. Was it difficult to estimate the distance? (If he answered yes, he was asked why.)
2. Show with a diagram the way the light moved.
3. Did you try to find some method of your own, so that you could make your judgments more accurate?

Some of the representative answers [4] to the first question were:

> Darkness left no guide for distance.
> No set position from which to judge how far.
> Didn't know direction it would move.
> Lack of visible neighboring objects.
> No fixed point from which to judge distance.

These answers confirm Sherif's expectation in setting up the experiment—that there would be no very clear frame of refer-

[3] As can be noted in Fig. 3, this statement does not hold for a few subjects whose median or modal extent of illusory movement is very near one end of the distribution.

[4] Sherif, *op. cit.*, pp. 24-25.

ence, at least in the beginning of the experiment, within which to judge extent of movement.

The answers to the third question [4] actually constitute the major result of the experiment. Some of these were:

Compared with previous distance.
Judgments are all relative.
Compared successive judgments.
Approximated distance of spot from me, and used that.
First estimate as standard.
Thought of using radium dial of watch for judging distances.

The quantitative results of Fig. 3 showed that by some means or other the subjects must have set up crude standards by which to judge extents of movement or they could not have been so consistent. The answers to the third question, we see from the examples just given, show what the subjects actually did. They tried to put their judgments within some sort of framework—movement was judged with respect to the movement just previous, or with respect to the very first estimate, or in some way placed on a consistent basis, however vague and unsatisfactory this might actually be. Apparently a frame of reference is so necessary that when it is lacking the subject "creates" at least a minimum basis for judgment.

SHERIF'S GROUP EXPERIMENTS

We have omitted consideration of one of the most important factors in the development of a frame of reference for conduct. A very large proportion of our learning and of our judgments is *socially directed* or *socially influenced*. It therefore is of interest to determine how the presence of a group influences judgments as to movement of the light in our simple experimental situation.

In the lower half of Fig. 2 is shown the essential apparatus for doing the autokinetic experiment with small groups (two or three subjects). Each subject at the table (shown in Fig. 1) had before him a reaction key and a signal button. Any reaction key would start the automatic timer. The signal buttons

[4] Sherif, *op. cit.*, pp. 24-25.

were connected with different colored shielded lights on the experimenter's table. These lights were dim and could not be seen by the subjects. By this means, when reports were given, the experimenter could identify a subject speaking in the dark, if the latter held his finger on his signal button as he spoke.

The following instructions were given:

> When the room is completely dark, I shall give you the signal ready, and then show you a point of light. After a short time the light will start to move. As soon as you see it move, press the key. (Press it the moment you see the light move. Don't wait for the other persons.) A few seconds later the light will disappear. Then tell me the distance it moved. When you give your estimate, press the push-button. Try to make your estimates as accurate as possible.[5]

After the instructions had been given, the subjects were told that they could give their judgments in any order, and they could change the order from time to time, so that the same person did not always report first.

There were eight groups of two subjects each, and eight groups of three subjects each. Half the groups began with the individual situation (one sitting for each person) and then came together as a group. Each of the other groups began together and a later sitting came individually. The general plan of experimentation may be schematically presented as follows:

Starting with the Individual Situation

(Four groups of two, four groups of three)

Session:	I	II	III	IV
	Individual	Group	Group	Group

Starting with the Group Situation

(Four groups of two, four groups of three)

Session:	I	II	III	IV
	Group	Group	Group	Individual

The results of these experiments are summarized in Figs. 4 and 5. The conclusions to be drawn are fairly clear-cut. First, if subjects come into this strange situation for the first time

[5] M. Sherif, *op. cit.*, pp. 28-29.

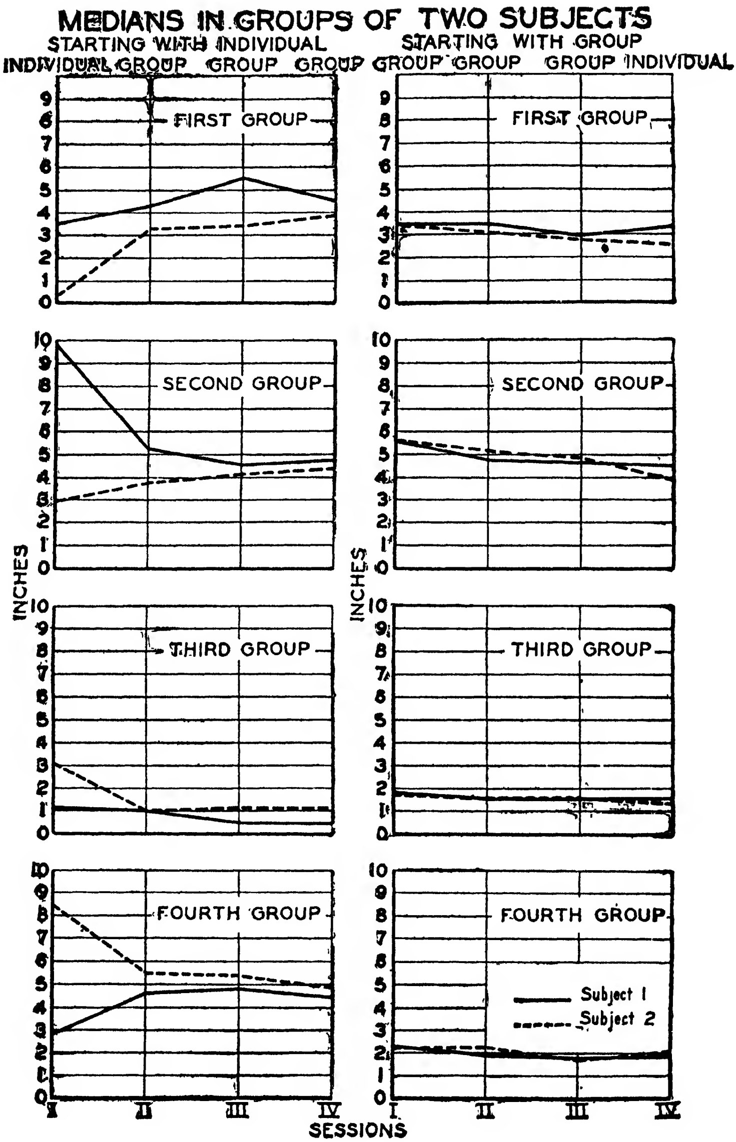

MEDIANS IN GROUPS OF TWO SUBJECTS
STARTING WITH INDIVIDUAL
STARTING WITH GROUP
INDIVIDUAL GROUP GROUP GROUP
GROUP GROUP GROUP INDIVIDUAL
FIRST GROUP
SECOND GROUP
THIRD GROUP
FOURTH GROUP
INCHES
Subject 1
Subject 2
I II III IV
SESSIONS

FIG. 4.

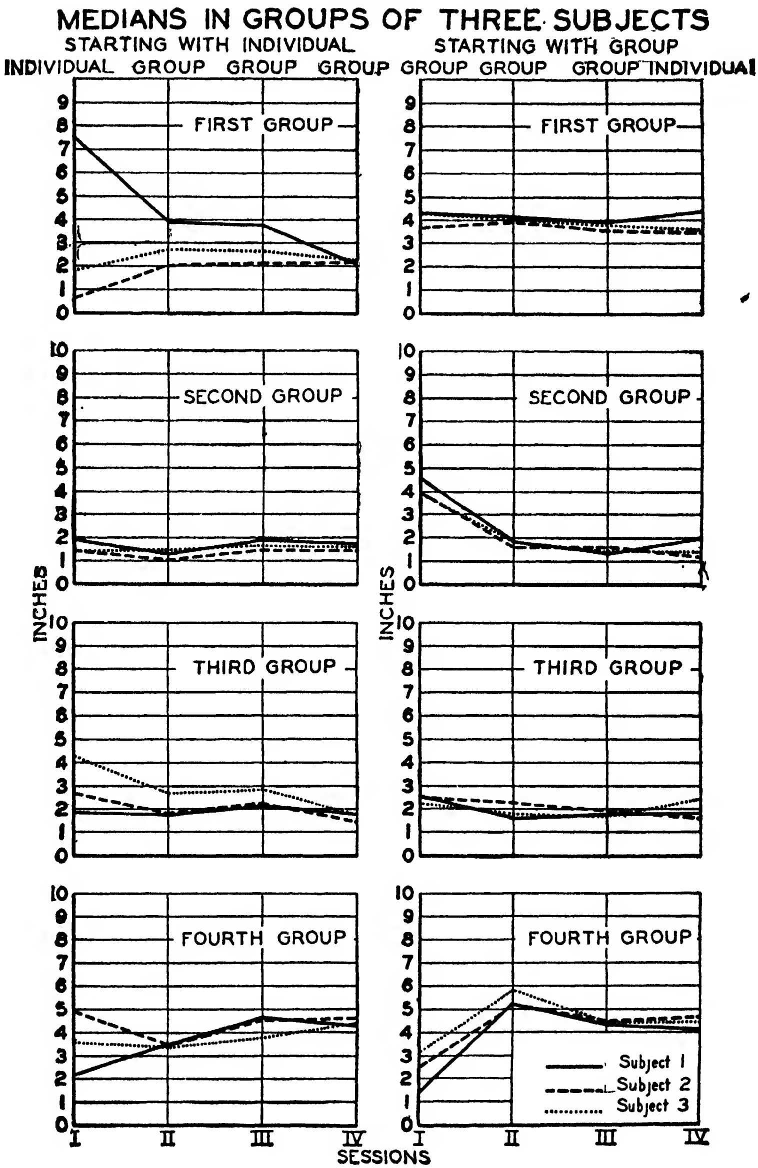

MEDIANS IN GROUPS OF THREE SUBJECTS
STARTING WITH INDIVIDUAL
STARTING WITH GROUP
INDIVIDUAL GROUP GROUP GROUP
GROUP GROUP GROUP INDIVIDUAL
FIRST GROUP
SECOND GROUP
THIRD GROUP
FOURTH GROUP
INCHES
SESSIONS
I
II
III
IV
Subject 1
Subject 2
Subject 3

FIG. 5.

singly, each person sets up for himself a general range and a standard extent of movement within that range. The standards tend to converge when the group comes together for successive meetings, producing a funnel-shaped graph. Second, if subjects begin the whole experiment as members of a group, a range and a standard characteristic of the *group* are set up in the first sessions; when these subjects later have individual sessions, they tend to stick to the standard set up in the group. Individual differences in response to an ambiguous perceptual situation thus are smaller in the group. Furthermore, while the effect of the group carries over into later individual sessions, the standards set up first in individual sessions give way in the presence of the group.

When subjects in the group experiments were asked whether they were influenced by judgments of the others in the group, only about one-fourth of the subjects replied affirmatively. The graphs of Figs. 4 and 5, of course, show a much greater influence than this—signifying probably that it is not necessary to be aware of an influence on one's judgment in order to respond to this influence, a point made previously in this chapter.

In the *group* experiments which Sherif conducted, then, an additional factor entered into the frame of reference of subjects estimating distance which the light moved. This was the reported judgment of other members of the group. One may be inclined to marvel at the fact that the mere opinions of others should influence the extent of an illusion, which probably has its subtle physiological basis. Probably one should not marvel: only in a rather vague and ambiguous situation, such as the autokinetic experiment in a completely darkened room, would the group influence perception so markedly and consistently. Visual perception ordinarily is governed quite largely by the external stimulating situation, *if* it is at all clear what is supposed to be perceived. But where the influence of the physical situation does not clearly "dictate" the appropriate response, as in the autokinetic experiment or in the learning of ethical codes, other social factors are quite certain to have a large influence on the behavior of the individual. The

group in this case contributes to the frame of reference of each of its members.

Summary and conclusions. The frame of reference is generally thought of as the background of stimulation influencing response to outstanding stimuli. Examples of how the frame of reference operates in various situations are fairly easy to compile. It has been shown, furthermore, how a framework may come to operate even where it is in the beginning relatively ineffective. Experiments based upon the autokinetic phenomenon show that the subject, where necessary, develops his own frame of reference, by way of memory for his previous judgments, or through the use of standards based on some background aspect of the present visual situation, or by means of the reported judgments of members of his group. This statement leads to a major conclusion: depending upon the kind of situation in which a person finds himself, either of the two major parts of the frame of reference, namely, stimuli to the sense-organs or ideas, attitudes, etc., based upon previous experience, may be the dominant part of the background stimulation influencing responses to the outstanding stimuli. Both, of course, may be and usually are effective at the same time.

CHAPTER 3

Sensitivity of Hearing

THE AVERAGE person can perceive tones produced by sound waves varying in frequency from about 16 to 20,000 cycles (or double vibrations) per second. When a person becomes partially deaf or hard of hearing we often find that he can still hear very faint tones of certain pitches, only moderately loud tones at other pitches, and none at all in still other parts of the pitch range. It therefore becomes important to know just how much his hearing has been impaired in each part of the pitch range,[1] and this is best described by comparing the faintest sounds he can hear (if any) with what the normal person can hear.

Our first experiment describes measurements of the faintest tones which can be heard by adults of normal hearing. Such measurements of the lower limits of perceptibility are called absolute thresholds or limens.[2] The *absolute limen* is defined as the minimum degree of a stimulus necessary for perception of

[1] Pitch is determined mainly by the frequency of vibration of a tone. A tone coming from a stimulus source which is vibrating very rapidly (e.g., a violin string) is higher in pitch than a sound coming from a source vibrating more slowly (e.g., a bass viol string). *Loudness* is governed chiefly by the amplitude of vibration of a tone. It should be noted that pitch and loudness refer to the *experience* of hearing; frequency and intensity (amplitude) refer to aspects of the physical stimulus. In the experiments to be discussed here the stimulus has been varied in order to determine what effect this has upon experience. Therefore our measures of sensitivity will be stated in stimulus units but will relate to experience, since, for example, a change in frequency is significant if it produces a perceptible change in pitch.

[2] The phrase "terminal threshold" is also used here. It should be remembered that tones may be of frequencies too high or low to be audible just as lights may be of wave lengths too long or short to be visible. In these cases we may speak of both upper and lower absolute thresholds.

its presence. In the present case, the absolute limen is measured in terms of the intensity necessary in order for a tone of any given pitch to become just audible.

The absolute limen at various pitch levels. In order to determine the absolute intensity limen for hearing, as this limen has been defined above, an apparatus must be used which will produce a tone of any desired frequency and which also is capable of varying the intensity of this tone, from below the absolute limen to any desired level above it. Such an instrument was used by Wegel. The absolute limens of ten subjects with normal hearing were measured for tones of frequencies from 35 to 1000 double vibrations per second.[3] The experimental frequencies were five double vibrations apart, between 35 and 100 double vibrations, and were spaced at larger intervals between 100 and 1000. Wegel extends his own measurements of sensitivity to frequencies above 1000 by including data secured on seven subjects by W. A. Munson. Since the results obtained by the two investigators were similar, we shall combine them, as Wegel did, into one smoothed curve showing average absolute limens at the various frequency levels.

It is evident in Fig. 6 that the normal person can hear very high or very low tones only if their intensity is much greater than that necessary for tones toward the middle of the pitch range. A tone within the range of about 1000 to 2500 double vibrations per second can be heard at a lower intensity than the minimum for all other pitches. It might be noted that the majority of voices and orchestral instruments produce sounds within the frequency range of 80 to 1500 double vibrations per second, although, of course, we can hear pitches somewhat lower and much higher than these.

Sensitivity to changes of pitch. A second kind of limen is the *difference limen,* which is the minimal increment (in fre-

[3] Wegel, R. L., "Physical Data and Physiology of Excitation of the Auditory Nerve," *Annals of Otol., Rhin., and Laryng., 41,* 1932, 740-779 (esp. pp. 769-771). The method used in presenting stimuli to the subject is not described. Presumably the tone being tested was presented below the threshold and gradually made more intense until the subject could just detect it.

quency in the present case) in a stimulus needed to produce a just-noticeable difference in experience. It thus is measured in units applicable to the stimulus in question, for example, some number of cycles (or double vibrations) per second. The *relative difference limen* is the ratio of the difference limen to the value of the stimulus to which it is added. It actually is the

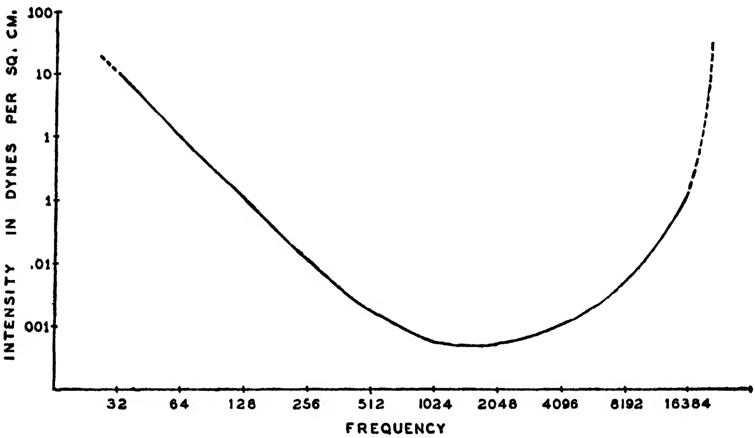


FIG. 6. Curve from Wegel showing absolute limens at frequencies ranging from 32 to 16, 384 double vibrations per second.

percentage of change in the stimulus which is necessary to produce a change in experience.

The relative difference limen will be applied in the following way. In the case of perception of changes in pitch, the subject hears a·certain tone of, say, 256 double vibrations per second (about middle C on the piano). At intervals the frequency of this tone is varied some very small, and at first imperceptible, amount. The amount of variation is increased gradually until the subject can just detect the inconstancy of the tone he hears. The difference limen will then be the number of double vibrations per second by which the standard tone had to be varied before the change was detected, and the relative difference limen will be the ratio of this number of double vibrations per second to 256, the frequency of the standard stimulus tone.

The smaller the relative difference limen, the more acute is the subject's ability to perceive changes in pitch. We can measure ability to perceive changes in loudness with the same general type of procedure.

Shower and Biddulph[4] have studied the perception of changes in pitch of pure tones having frequencies over a very wide range. A tone of any desired frequency within this range could be electrically produced in a special apparatus, and transmitted to a receiver placed over one of the subject's ears. This tone could be varied in frequency from the standard, about twice per second, and the size of the change in frequency gradually made greater until the subject could just perceive the fluctuation. To him the difference appeared in the form of a slight pulsation or lack of evenness in a single tone.

In order to control the factor of loudness in these measurements, the absolute intensity limen of the tone having the chosen pitch was first determined with pitch constant, and then the intensity of the tone increased a standard amount above the limen. In the data selected for discussion here the increase in intensity above the absolute limen was the same for tones of all pitches. In this way measurements of relative difference limens for the hearing of pitch could be based upon tones which were approximately the same distance above the threshold of audibility with respect to loudness.

In Fig. 7 are shown the results for the ten ears of five men between the ages of 20 and 30 years. At least ten, and in some cases as many as twenty, observations were taken to determine the difference limen for a given tone for a single individual. Thus, the points on the curve shown in Fig. 7 are averages based upon a minimum of 100 determinations and usually upon a larger number. Fig. 7 is based on experimental tones which were 40 decibels above the absolute intensity limen, the decibel being a standard unit indicating the relative intensity of a tone. We may get something of an idea as to how far above threshold these tones were by noting that various investigators have

[4] Shower, E. G., and Biddulph, R., "Differential Pitch Sensitivity of the Ear," *J. Acoust. Soc. Amer.*, *3*, 1931-2, 275-287.

found approximately a 40-decibel sound level in the following cases: a very quiet radio running in a home; a lectūre room containing 200 students who were as quiet as possible; a single typewriter being operated in a rather small office.

The outstanding fact shown by this curve is that the ear is less sensitive to changes in pitch of low tones than to changes in the higher tones. The ability to distinguish pitches increases

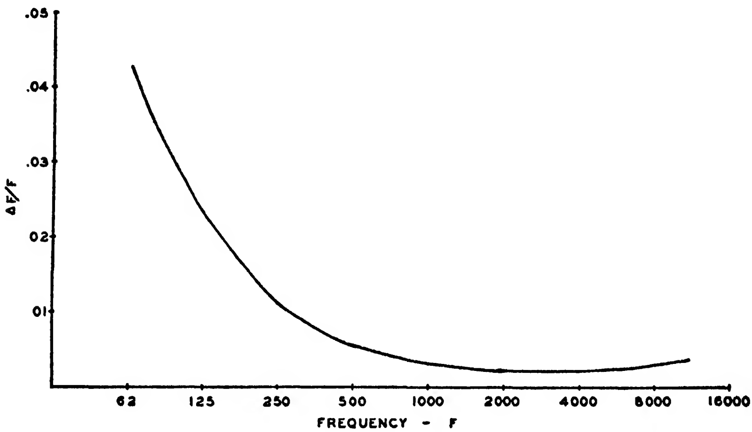


FIG. 7.

rapidly (the relative difference limen becomes smaller) up to frequencies of 500 double vibrations per second, and remains rather constant through the higher range, probably decreasing somewhat at the extreme upper pitches. At about 2000 to 2500 double vibrations per second the relative difference limen reaches its lowest value; i.e., pitch discrimination is keenest in this range.

Sensitivity to changes in loudness. We have seen that responsiveness to pitch changes varies according to the part of the audible pitch range in which the measurement is made. The experiment now to be reported deals with sensitivity to changes in loudness. In this experiment, performed by Riesz,[5] approxi-

[5] Riesz, R. R., "Differential Intensity Sensitivity of the Ear for Pure Tones," *Phys. Rev., 31,* 1928, 867-875.

mately the same procedure was used to determine the relative difference limen as was used by Shower and Biddulph. That is, a special electrical apparatus produced a tone of any frequency within a very wide range. In contrast with the Shower and Biddulph apparatus, it also produced a second tone of a frequency which was just enough different from that of the first to beat with it about three times per second.[6] The second tone was "added" to the first, beginning below the absolute limen of the second, and gradually increased in intensity until the subject could perceive that he was no longer hearing an even tone. The difference appeared to the subject as a beat effect, but the beats could not be reported until the intensity of the second tone was raised to a level where it was perceptible in the presence of the tone of standard intensity. Even then it was perceived not as a second tone but as a variation in a single tone.

In their experiment, Shower and Biddulph found the threshold of audibility for each tone used, whatever its pitch, and then raised it to 40 decibels above the threshold before varying it in pitch for purposes of the experiment. Riesz, on the other hand, used each experimental pitch at seven different levels above the absolute limen. In each case he added the beating tone below its threshold of audibility and raised it until the experimental tone, whatever its intensity level, was perceived to fluctuate in loudness.

Twelve male subjects were used by Riesz. In determining the relative difference limen in intensity, the standard tone (e.g., one with a frequency of 256 double vibrations per second at an intensity of 20 decibels above the absolute limen) would be presented and then the beating tone added. The latter was increased in intensity by some small amount and the subject

[6] If two tones of slightly different frequencies are sounded simultaneously, a single tone will be heard which alternately swells and decreases slightly in loudness. The number of swells or beats per second is determined by the difference in the number of double vibrations per second between the two frequencies. Thus a tone of 256 double vibrations per second from a tuning fork will beat three times a second with an adjacent fork of frequency 259.

asked to report whether he heard an even tone or a pulsating tone. The hearing of beats was taken as indicating that the difference limen, as measured by the intensity of the added tone, had been reached. About 20 judgments were required of the

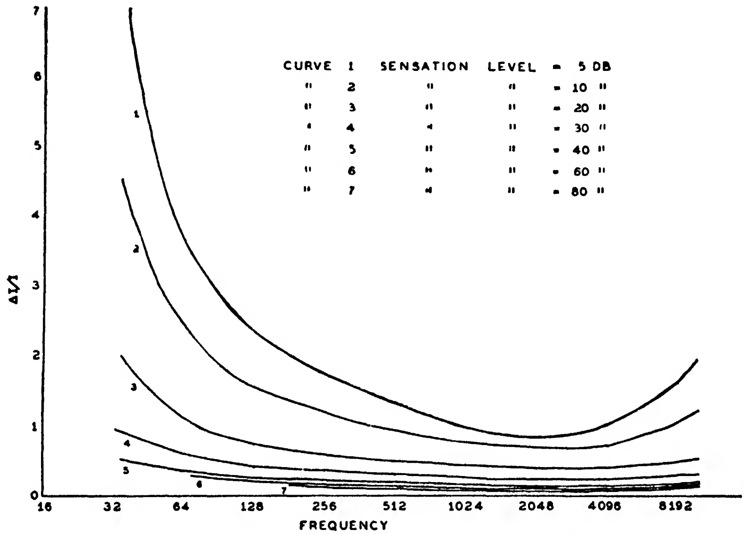


Fig. 8. Differential intensity sensitivity of the ear for pure tones.

subject in order to determine with assurance the exact size of limen for the tone being experimented with at that moment. He would then be given some new tone and the whole procedure repeated.

The results of Riesz' experiment are shown in Fig. 8. The graph there shown may be read as follows: the average relative difference limen for a tone of 256 double vibrations per second (about middle C on the piano), when it is at intensity level 2 (10 decibels above threshold), is about 1.2 per cent. Some approximations to sound levels in various situations are: loud radio music in the home, 80 decibels; very noisy restaurant, 70; lecture room, 200 students making no conscious effort to be quiet, or a noisy department store, 60; moderate restaurant clatter, 50; turning the page of a newspaper, 30; rustle of

leaves in a gentle breeze, 20; outdoor minimum in country, 10. In Riesz' work, it should be remembered, the experimental tones were rated in terms of distances *above* the absolute limen.

Figure 8 indicates that for tones 60 decibels and more above threshold, sensitivity to *changes* in intensity is about equal throughout the pitch range. Below 60 decibels, however, and particularly for very soft tones, perception of changes in intensity is much poorer in the lower pitch ranges, and probably also in the upper. The greatest sensitivity to changes in loudness is in the region of 2500 double vibrations per second, for all the levels of loudness studied by Riesz.

Summary. The absolute intensity limen is smallest (sensitivity is greatest) for tones in the lower center portion of the audible pitch range. At the upper and lower extremes of the pitch range a sound must be much more intense before it can be perceived at all.

The relative difference limen has been used as an indication of sensitivity to change in auditory stimuli. The studies here reported indicate that, with intensity held constant, the smallest relative difference limens for pitch changes occur in the neighborhood of 2000-2500 double vibrations per second. The measurements of sensitivity to changes in intensity indicate that for tones which are relatively loud to begin with, less increase relatively is necessary for perception of a difference than for tones which are soft to begin with. Furthermore, the greatest sensitivity to intensity changes is for tones around 2500 double vibrations per second in frequency.

The three studies taken together indicate that both absolute and relative sensitivity are less at the upper and lower perceptible extremes of pitch and loudness than in the middle range of stimulation.

CHAPTER 4

The Startle Pattern[1]

A COMPLEX motor reaction which is called startle is one of the first and most characteristic responses to many emotion-arousing situations (e.g., those capable of causing fear). Although everyone knows approximately what it means to be startled, this kind of behavior has not been fully investigated until recently. For one reason, the startle response takes place so quickly that the person who is startled cannot report adequately on what he does or feels during these few moments. Secondly, objective techniques for studying the startle reaction have not been sufficiently sensitive or accurate to record many of the important details. It has been easier to study major changes in bodily tension, glandular function, etc., which occur more slowly and can be recorded more satisfactorily. But with the increasing perfection of high speed photography it has become possible to collect relatively permanent, detailed, and accurate records of motor responses taking place in a small fraction of a second. Equipped with such a technique we can now study such problems as the following: the usual pattern of the startle response in normal adults, the time-sequence in which the components of the startle reaction occur, the kinds of factors which modify the reaction, and the types of stimuli adequate to call it out.

Procedure. If a motion-picture camera is run at such a rate as to expose eight frames of film per second, and the film thus secured is projected at a normal rate, e.g., 16 frames per second, the activities thus recorded and treated are crowded into a shorter space of time and the apparent rate of their occurrence is thereby speeded up. Conversely, if the camera taking

[1] Adapted from Landis, C., and Hunt, W., *The Startle Pattern,* Farrar and Rinehart, New York, 1939, xii + 168 pp.

the pictures is run at a rate of 64 or more exposures per second but the film is projected at 16 frames per second, the activities pictured will be spread out over a longer interval of time and the apparent rate of their occurrence will be slower. This basic principle of "slow motion" photography (as it is commonly demonstrated in news reels of athletic events) is the technique here used to study the startle response. Motion pictures taken at high rates of speed and later shown at normal speeds "slow down" the activity being observed and permit experimenters to study it in fuller detail.

In a typical experiment, a college student dressed in a bathing suit may be asked to stand before a black cloth back-drop, under bright illumination. A clear full-face or profile moving picture may thus be made of his behavior. A stimulus (e.g., the sound of a revolver being fired near him) is presented, and the high-speed cameras record his every movement and facial expression thereafter. When pictures are taken at 64 or more exposures per second, an electric stop-clock showing 1/100 second intervals is placed in the camera field, so that the behavior of the subject is automatically recorded in terms of such units. When superspeed cameras are used, at rates of 300 to 3000 exposures per second, a special timing device is necessary. This consists of a revolving dial which is fitted into the camera base; the image of this dial is projected by a prism system upon the edge of the film so that intervals of time smaller than those which could be read adequately from the electric stop-clock become usable. All the timing devices are synchronized with the stimulus gun so that the discharge of the revolver starts the stop-clock or the special revolving dial.

In scoring the startle responses a projector is used which enables the experimenters to run the film backward or forward, or to keep a single frame in place for detailed consideration. At least two, and usually three or four, persons judge each film. In the main, their task is to state whether or not a certain type of movement is present in the picture, although they occasionally give judgments as to the intensity or amount of response. Since but one element in the person's movements is scored at a

time, it is necessary to re-show the film to the judges as many as twenty or more times, with numerous repetitions for preliminary familiarization and other general impressions.

The normal response. As a reference experiment, the startle reactions of well over 100 normal adults have been studied, using the firing of .22 and .32 caliber revolvers as stimuli, with cameras running from 64 up to 2200 exposures per second in various experiments. These reactions fall into a definite pattern, which is much the same in a given person from time to time and much the same from person to person. The pattern of response includes the following: (1) blinking of the eyes, (2) head movement forward, (3) a characteristic facial expression, (4) raising and drawing forward of the shoulders, (5) movement of the upper arms away from the body, (6) bending of the elbows, (7) inward turning of the lower arms, (8) flexion of the fingers, (9) forward movement of the trunk, (10) contraction of the abdomen, and (11) bending of the knees (see Fig. 9). Because the appearance of this general pattern is predictable, and particularly because certain parts of it are quite stable in spite of various modifying conditions, as will be described later, the startle response just described may be considered basic and universal among normal adults.

The characteristic facial expression mentioned above may be analyzed more carefully. As already mentioned, the most noticeable feature in the facial reaction is the immediate closing of the eyes. Then a widening of the mouth occurs, as though in a grin. However, only occasionally are the teeth really bared during this part of the response. As the head and neck come forward, the chin sometimes is tilted up so that the face is still directed straight ahead. Neck muscles characteristically stand out under tension and in rare instances a twitching movement in the scalp and ears can be noticed.

In mild startle reactions the change in facial expression is the more noticeable part of the total pattern; in strong reactions, however, the bodily reaction becomes the more obvious. This bodily reaction (as may be seen in Fig. 9) can be de-

scribed in a general way as a flexion of the body which is something like a protective contraction.

Although we have stated that the *general* pattern of reaction in startle is consistent in appearance, some of the elements of

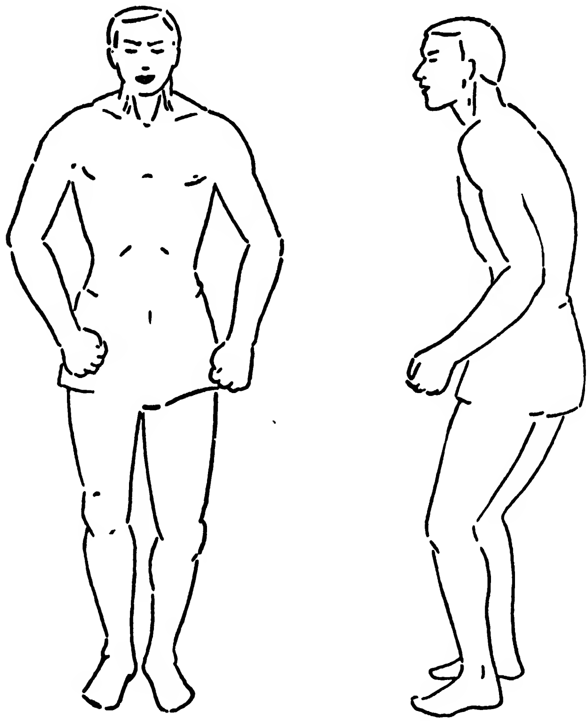


FIG. 9. Schematic representation of the bodily pattern. (Taken from Landis, C., and Hunt, W.: *The Startle Pattern,* p. 22).

the pattern previously listed may not occur at all in certain persons or may not appear in every response by a given person. Consistency is thus a relative word in this context. The eye-blink seems *always* to occur in the normal adult, and although in exceptional instances *only* this may appear, most persons show some of the characteristic head movements and also some of the characteristic bodily reactions. Even where ordinary observation seems occasionally to find a type of movement op-

posed to the general flexion pattern (as in the case of a backward arching of the trunk or an extension of the fingers), closer observation or sensitive recording reveals in the majority of cases that flexion occurred first and the apparently exceptional kind of movement was a secondary phenomenon.

The startle response occurs in a surprisingly short period of time. It usually is completed in less than half a second after the stimulus occurs. The simultaneous blinks of the two eyes are the first element of the pattern to appear; the average time for this movement to begin, in 45 startle responses from 11 persons, was .04 second. The time from the beginning of the response to complete closure was about .015 second; as might be expected, the behavior after this point was much more variable, with some subjects opening their eyes immediately, others slowly, and others not for some moments. Although it is more difficult to be exact about the time of occurrence of other elements in the movement pattern, approximate averages for the beginning of a few responses were: widening of the mouth, .069 second; head movement, .083 second; movement in the neck muscles, .088 second. Although there are large individual differences in the absolute reaction times of these various responses, the relative temporal order is always: eyelid closure, mouth movement, head and neck movement at approximately the same time, and then movements of the remainder of the body. With reference to parts of the body, the startle pattern begins in the eye-blink; the next movements to occur are in sequence *downward* over the remainder of the body, with the movements of trunk, arms, etc., coming after the responses of the upper body.

With the above discussion as a reference point, we can now attempt to answer some of the questions as to the characteristics of the startle response under conditions other than the relatively simple and standard ones involved thus far.

Startle as dependent upon location of the auditory stimulus. One may at first think that the general flexion response to the sound of a gun is simply a retreat in a direction away from the possibly harmful stimulus. In the films it seems to

be clear, however, that the startle response is independent of the location of the sound source. The shot may come from above the subject's head, below his knees, in front of him, at either side or behind, and the pattern remains unchanged.

Knowledge of the stimulus. It cannot be predicted in advance for a particular person just how much influence his knowledge about the gun, expectation of the time when it will be fired, etc., will have on his startle response. If subjects are given this information about the stimulus, some of them will, it is true, show a reduced amount of response, although the absence of surprise as such will not prevent the reaction altogether. Other subjects show the usual complete or intense response although they know when the gun is to be fired. This holds true even when such a subject himself fires the gun—the eye-blink and usually head movement can be detected. In subjects who show lessened response when they are not surprised by the gun, it is possible to "step up" the reaction by using a more intense stimulus, e.g., a .32 caliber gun instead of one of .22 caliber. It thus seems possible that a startle stimulus must be described in terms of at least two factors: surprisingness and intensity. If the element of surprise is lacking, the sheer intensity of the stimulus may still call out the startle response.

Adaptation. When shots are fired every one or two minutes, the subject may show signs of adaptation,[2] or reduced response. There are wide individual differences in the rates at which such adaptation may occur, but in no experiment thus far has the eye-blink (ordinarily the most dependable part of the general movement pattern) disappeared, and usually head movements continue to occur. In other words, the components of the startle reaction which appear first after the stimulus is given and which are most generally found among normal subjects are the ones which are least likely to disappear under repeated presentation of the stimulus. It should be said that some persons show little or no adaptation—one such subject gave the complete startle pattern to 17 successive shots spaced a minute apart.

[2] Learning "not to respond," as in this part of the experiment, is usually called negative adaptation.

Further interesting proof of the failure of the startle response to become completely adapted (although many kinds of responses show great adaptation with repetition of the stimulus) is furnished by a study of trained marksmen. These men were officers in the New York police department, were members of various pistol teams, and had administrative or instructional duties on the pistol range. Seven men were included in the study; their duties had for years subjected them to stimulation on the pistol range, listening to both the shooting by others and the sounds of their own guns during target practice. These men provide a test of negative adaptation (and effect of the knowledge about the stimulus as well) based on much longer experience with guns than could be given in a laboratory experiment. Motion pictures were taken on the regular pistol range as each man fired a .38 caliber pistol five times in succession. Because of the lighting and other handicaps attendant upon photography under these circumstances, it was possible to get satisfactory records on only head movement and facial expression. The results were clear-cut: every man showed an eye-blink to every shot; six of the seven showed head movement as well (and the seventh did upon one occasion); most of them showed mild facial distortion. Thus, after a long opportunity for negative adaptation, certain movement elements of the general startle response continue to appear. Surprisingly enough, the men were not conscious of the blinking, head movements, and facial expressions, and some of them even denied the possibility of such behavior—but the films convincingly demonstrated that they "made faces" despite their long experience.

The above facts concerning the influence (or lack of influence) of knowledge about the location of the stimulus, the time when it will occur, and continued experience with it, now permit us to make the general statement that the startle response is mainly outside the influence of voluntary control.

Inhibition and facilitation. The question at once arises as to just how completely involuntary the response is, because in the work thus far reported, the subject was not instructed to

modify his responses deliberately. Evidence from experiments in which the subject did try to suppress (inhibit) the reaction or increase (facilitate) it leads us to think that the *amount* of response can be modified somewhat if the subject so intends, but that it never disappears completely or increases markedly.

The experiments where subjects tried to inhibit the response do not require discussion, because these attempts never eliminated the startle. But where the subject was asked to get set and facilitate the response by actually jumping at the sound of the shot, some interesting kinds of behavior were observed. Superficially the response in this case seemed to be increased in amount, with many atypical movements added to it. The superspeed cameras show, however, that the primary pattern occurs just as usual after the shot, and *then,* a few thousandths of a second later, there occurs a voluntary duplication of the response. This secondary response is apparently the one in which the subject has set himself to produce; it is superimposed upon the typical, quick response, and actually is not a very good imitation, since it includes atypical elements of movement. Occasionally the secondary response follows so closely after the primary, in the facilitation experiments, that it is very difficult to separate the two in analyzing the films. With the exception of such occurrences, it seems to be fairly clear that the voluntary attempt to facilitate the startle response results in little change in the primary or true reaction, but does result in the addition of similar voluntarily controlled behavior following quickly after the primary pattern.

Other stimuli. Four stimuli other than the firing of a revolver were tested for their effectiveness. Fifteen subjects were told to dress in one-piece bathing suits for the experiment. The stimuli used were: a jet of ice water between the shoulder blades, an electric shock delivered to the right hand, a pin-prick in the right thigh, and a discharge of two large magnesium-foil photoflash lamps. The jet of ice water was projected through a small hole in the back-drop before which the subjects stood. The shock was given through copper electrodes which were attached to the hands of the subjects under the pretext that

the experimenters were interested in measuring the electrical resistance of the skin. The pin was jabbed into the subject's thigh (the intensity of this stimulus naturally had to be kept within rather moderate limits!) when attention was momentarily diverted to the cameras. The flash-bulbs were discharged in the usual way, so that (as at a public gathering where news photographers are busy) the sudden flare of light caught the subject's eye. All of these stimuli are mainly effective in some modality other than hearing and, furthermore, resemble the sorts of things which often startle us in everyday life.

Of the total of 15 subjects, only one showed a complete startle response to the discharge of the flash-bulbs; two showed the eye-blink, head movement, and shoulder movement; three showed the eye-blink and head movement, and four showed merely the blink. The remaining five subjects gave no response at all. Although it is apparent that a sudden visual stimulus *can* call out the startle reaction, it does not seem to be as certain as an auditory stimulus.

The experiments with shock stimulation were more difficult to interpret, because the reaction in this case included some or many specific movements of the arm or hand shocked. When such movements occur, in response to a localized stimulus, the usual general startle pattern is to some degree modified by the localized behavior, and, also, the localized movements may interfere with the observation of the general response. In spite of these difficulties five of the 15 subjects seemed to show fairly complete startle reactions, and five more, parts of the pattern. The remaining five simply responded with movements limited to the hand and arm shocked. Shock, then, also is to be classed as capable of eliciting the startle pattern, but it, like visual stimulation, is probably not as effective as a sound.

Like the shock, the jab of a pin produced marked avoidance behavior which served to confuse the general startle reaction picture. Two of the subjects gave complete startle patterns, but included swaying the body to the left, away from the pin. Two more subjects gave the eye-blink. Four subjects simply showed avoidance and none of the usual startle pattern, and the re-

mainder did not respond at all. Although the latter fact is probably an indication that the pin was not used vigorously enough, it seems probable that this kind of stimulation *can* call out the startle reaction but not so frequently or so completely.

The jet of cold water can be described conservatively as being very effective. Eight of the 15 subjects gave complete or partial startle responses, three gave immediate avoidance responses nothing like startle, and the remainder of the subjects showed no (immediate!) response.

None of these stimuli, at least as employed thus far, seem to be as effective in calling out the usual startle pattern of movement as is discharge of the revolver. It is quite possible, of course, that none of the other stimuli tested were as intense as the auditory stimulus. Although it is difficult to make comparisons as to intensity of stimuli in different modalities, the subjects used here were very definite in saying that the gunshot was the most intense stimulation they experienced. From this point of view, then, it seems that one should be somewhat cautious in rating the effectiveness of auditory as compared with visual and tactual stimuli.

Furthermore, the startle pattern may be described as a *general* pattern of movement. Auditory stimuli, which are not so easily localized by the subject as are tactual and visual stimuli, might be *expected* to call out the general type of reaction. Localizable stimuli such as a pin-prick would be expected to call out suitable avoidance responses and, less effectively, additional elements of the general response.

Secondary behavior. In addition to the immediate, involuntary muscular reaction which we have described, certain other behavior also occurs in response to a startle stimulus.

Secondary reactions to the startle stimulus (those which are later in occurrence) may be classified in four main categories: curiosity, fear, annoyance, and overflow effects. Each of these, following upon a first general reaction, may be a rather complicated kind of behavior, in comparison with the primary startle pattern. The secondary reactions are much more variable from time to time and from subject to subject, and are

apparently more subject to voluntary control of the individual.

Evidences of curiosity following the primary reactions occur when the subject looks at or pays attention to the stimulus source, or other parts of the stimulus situation, or to the purpose of the experiment. He may turn his head or body or raise his eyebrows when he does this, merely appearing puzzled. Running away after the gun is fired, or defensive reactions such as covering the face or ears with the hands, are signs of fear. Perhaps because the subjects realize they are doing an experiment, or wish to be cooperative, angry attack on the experimenter or apparatus does not occur, but lesser signs of annoyance do often appear. Gestures, facial expressions, etc., may indicate irritation with the whole proceeding. The category of overflow effects includes the instances where the secondary behavior does not seem to be rational or directed but is simply a release of tension not discharged in the primary muscular reaction. Such things as changing posture, giggling nervously, smiling, or making inconsequential remarks are thus classified.

Many of these kinds of secondary behavior are simply verbal, or verbal additions to the motor behavior. The subject may say that he is frightened, or annoyed, or show a curiosity as to what is going on. He may be perfectly satisfied to express himself thus, and to show few or no other signs of secondary response.

This classification of responses is only tentative. A few other kinds of secondary behavior are not as yet satisfactorily identified and further experiments may show the need for somewhat different categories.

Summary. High speed photography has shown that in normal adults the typical startle pattern (of response to the sound of a gun) includes a very quick *eye-blink,* followed closely by the assuming of a *typical facial expression, movements of the head and shoulders,* and then *movements of other parts of the body.* The general picture is one of flexion or contraction of the body, as though in self-protection.

The location of the stimulus may be changed, or the subject may be warned in advance of what is coming, or allowed to

shoot the gun himself, *without greatly changing the response.* Repeated stimulations may be given, or the subject may try to inhibit or facilitate the response, and, although the amount of response may be changed somewhat (or, in some cases, a good deal) *the basic pattern of response stays much the same.*

Stimuli other than the firing of a gun (e.g., a sudden jet of ice water or a sudden flash of light) seem capable of calling out the startle response, *but not so definitely as does the sound.*

The main kinds of behavior which are temporally secondary to the startle pattern itself include curiosity, fear, annoyance, and overflow effects.

The fact that every normal adult shows the relatively predictable behavioral pattern of startle is significant. It adds to our understanding of what goes on in situations where suddenness and intensity of stimulation are characteristic antecedents of emotion.

The significant additions and corrections which this experimental method makes to our knowledge of the startle pattern suggest the value of similar ultra slow motion photography studies for other behavior, such as high speed discriminative reactions.

CHAPTER 5

The Nature of Mechanical Ability [1]

A MAJOR gain was made in our understanding of the nature of human abilities when it was realized that such terms as "general intelligence," "mechanical ability," and "musical ability" really represented *groups* of abilities which could be analyzed and measured in detail. In each of the three fields mentioned, the question has been: What are the various kinds of human abilities in this area, and how are they related to one another? Since the analyses of intellectual and musical abilities are reported in other chapters (20, 30), we shall limit our present discussion to the study of mechanical abilities.

If there were one kind of ability underlying success in all human activities we should call that ability a *general factor.* That is, the amount of this ability possessed by a given person would be a factor determining his achievement in *all* kinds of performances. If there were certain abilities each of which underlay success within a limited range of human activities, e.g., mechanical skills, we should call those *group factors,* and the amount of each ability possessed by a given person would be a factor underlying success within a certain group of activities but not in others. Such group factors might be significant for *all* mechanical skills (considered as a single rather large area of activities) or for only a smaller number of them. For example, there might be a group factor for the operations involved in mechanical drawing, and another for lathe operation. Thus we may speak of *broad* group factors—those which are important in many skills, or *narrow* group factors—those which are important in a smaller number of skills. In addition to the

[1] This chapter is based upon and is an expansion of: Harrell, Willard, "A Factor Analysis of Mechanical Ability Tests," *Psychometrika, 5,* 1940, 17-33.

possible general and group factors, we have long known that there are many *specific factors*.[2] For example, how much one knows about a "trick of a trade" may determine how well one does in a given task. The factor is specific to one kind of performance only and does not influence success in any other. Obviously there is not much difference between a specific factor and a very narrow group factor which is important in only two or three restricted skills.

Analyses to determine the number and kind of abilities important in the mechanical field have a very practical goal. This is the analysis and measurement of the abilities possessed by people who are successful in certain specific skills, such as the operation of various machines, or the assembling of apparatus, or manipulation of certain kinds of tools, to aid in the selection of applicants for these tasks. This analysis may possibly show that excellence in any one of these operations is related to all three of the kinds of factors mentioned previously. That is, lathe operation might be influenced by (1) the amount of some *general* ability possessed by the workman, (2) the amount of some ability important in all sorts of lathe operations but not in apparatus assembly (a *group* factor), and (3) the amount of some abilities which are each *specific* to various single operations he performs and unimportant in other lathe operations. Several group or specific factors thus might overlap in any complex mechanical operation.

Research into the nature of mechanical abilities has not as yet been very successful in analyzing out factors in the manner just mentioned. It is true that tests measuring many different kinds of mechanical ability have been used in attempts to select workers for various jobs. On the other hand it has never been very clear just what abilities each test measured. Therefore the choice of tests with which to attempt to select lathe opera-

[2] An ability is called a general, group, or specific factor insofar as the amount of it present in a given person is related to degree of success in all activities, a group of activities, or only one activity. Although we may refer, for example, to "group factors" without mentioning the word abilities, the implication of the term "group factors" always is: "an ability which is a factor in success in a group of activities."

tors or other workmen has been made mainly on the basis of trial-and-error by the experimenter.

The problem. In the experiment to be reported here the practical problem of worker selection has been postponed temporarily. Instead, a large number of mechanical tests and ratings have been made on a group of subjects. We shall attempt to find out in general *just what mechanical abilities there are* and how they are interrelated. After we have learned what some of the mechanical abilities are we shall be more nearly ready to study which of these are important in assembling or packing or other factory and shop operations.

The setting of the experiment. A total of 37 tests were given to 91 Georgia cotton mill machine fixers, who ranged in age from 19 to 51, with an average of 30.5 years. The job which these men hold requires the diagnosing of machine difficulties, general repairing, and making replacements of parts. They came from mills owned by three different companies.

Each subject spent about seven hours at the various tests required of him. This total time was divided into four different sessions, the first and last sessions being group testing, and the second and third individual. One of the three firms whose employees took part paid them for the time they spent as subjects. The two remaining firms simply permitted their employees to donate the necessary time. This differential treatment of subjects apparently had little or no effect on their motivation during the tests. The only difference between volunteer and paid groups was that the former were slightly less regular in attendance at the testing sessions.

The test battery. Although most of the variables included in this study were tests of mechanical ability developed by various investigators, certain other data concerning the subject were included. These were treated as though they were test scores. For example, the first three measures listed in Table 1 are the age of the subject (called youth), the amount of schooling he had, and the amount of mechanical experience (labeled inexperience). Strictly speaking, these are not tests of ability,

but they were included because they probably have some bearing upon test performance.

Variables 4 and 5 were two different types of score based on one test, the Minnesota spatial relations test[3] (Fig. 10). One score represented accuracy and one speed. The task of the subject was to choose wooden pieces corresponding to cut-outs in the test-board, and to insert these properly. This test was included in the battery because it seems to measure an essential in many shop performances—the ability to select parts, to fit

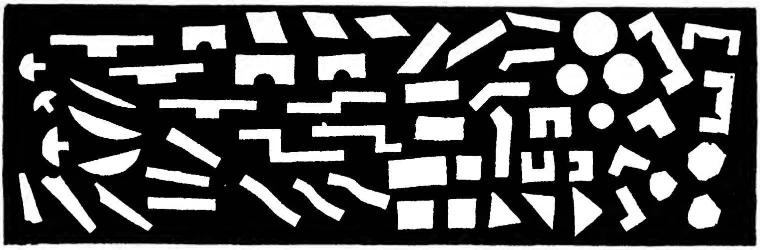


FIG. 10. The Minnesota spatial relations test.

them together, and to recognize a certain part as being the desired one even though it may be turned out of position when the subject first sees it. The two types of score were used to determine which was the better indicator of what the test measured.

The Minnesota assembly test was the sixth variable. The subject was given the task of putting together various simple bits of apparatus, with the aid of only a screw-driver. Parts for each object were placed together in a separate tray. In the sample test shown in Fig. 11, the objects are: an expansion nut, a hose pinch clamp, Hunt paper clip, clothespin, safety-link chain, wire bottle stopper, push button, bicycle bell, Corbin rim lock, and coin purse. The subject must in each case figure out how the parts fit together and then assemble them satisfactorily. Credit was given for assembling various objects within a cer-

[3] For this and other tests discussed in this section see Paterson, D. G., *et al., Minnesota Mechanical Ability Tests,* University of Minnesota Press, Minneapolis, 1930, pp. 586.

tain time, but more credit was given for the more difficult operations. The test performance required is quite similar to everyday practical shop work and as such seemed likely to measure some aspect of mechanical ability, in the sense in which we have used the term.

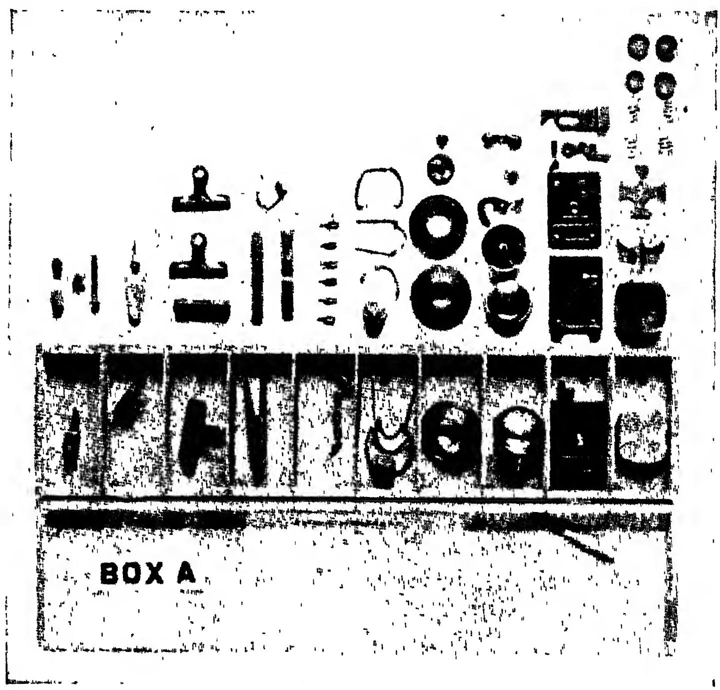
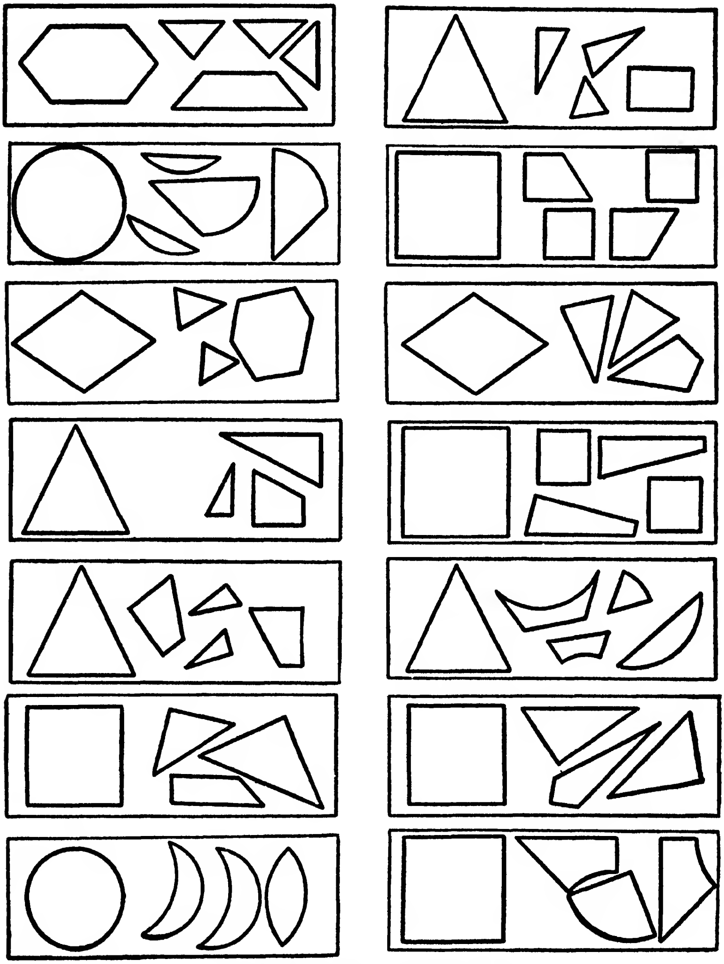


FIG. 11. The Minnesota assembly test.

Test 7 was the Minnesota pencil-and-paper form-board. A sample page of the test is shown in Fig. 12. The subject was required to show, by means of lines drawn with a pencil, how several small geometrical figures could be fitted exactly into a larger one. He was instructed to work as rapidly as possible. This test, like variables 4 and 5, seemingly necessitated an ability to perceive spatial relations, an essential in many mechanical operations.

An interest blank was the next variable in the battery. It was

thought that the likes and dislikes of the subject might be related to his performance on the tests. He was asked to rate his



Original dimensions, 8½ in. by 11 in.

FIG. 12. A sample page from the Minnesota pencil-and-paper form board test.

liking for each of a list of occupations, his likes and dislikes for various activities (including sports and hobbies, etc.). His score indicated his degree of interest in things mechanical.

The next two variables (9 and 10) were re-administrations of the Minnesota assembly test. The first was a speed test at assembling the objects previously assembled, and the second was a test at stripping or tearing down the same devices.

Tests 11, 12, and 13 utilized a pin-board. The subject was required to pick up inch-long brass pins from a tray and place them in a square board containing 100 holes in rows. In the first test the preferred hand was used, in the second, the non-preferred. In either case, the goal of the subject was to fill as many holes as possible within a time limit. For the third of these tests a board was used which had slightly larger holes. Both hands were used by the subject in a speed test at putting three pins in each hole. Nimbleness and sureness of hand movements, and eye-hand coordination, seemed to be required to do these simple tasks well. Test 14 was similar to the previous three, except that a larger board and wooden pegs were used. The subject was required to place the pegs in the board in certain patterns, e.g., place one, skip one hole, place one, etc., as quickly as possible.

Tests 15 and 16 called for a common shop operation. Twenty bolts, an inch in length, with nuts to match, were placed before the subject. In Test 15 he was required to strip the bolts of their nuts as quickly as possible. Test 16 then required that he replace the nuts, again as quickly as possible.

The next test also resembled a common shop procedure. A large number of wooden pegs had to be sorted out according to colors. This was also a speed test.

Test 18 was an interesting one, the so-called wiggly-block test.[4] The test-block is shown in Fig. 13. A rectangular wooden solid was so sawed that it consisted of nine irregularly shaped pieces. The task confronting the subject as he attempted to assemble the pieces was much like the solving of a Chinese puzzle. Presumably this test calls for an ability to imagine how various pieces fit together, in order to select the proper ones. Some dexterity in fitting them together quickly also

[4] From O'Connor, J., *Born That Way,* Williams and Wilkins, 1928, pp. 323.

seems necessary. The reasons for including this test are approximately the same as for tests 4 and 5.

Test 19 was similar to tests 15 and 16, except that 36 bolts of various lengths up to 1½ inches and of various diameters had to be fitted with nuts. Then the assembled object had to be placed in a hole of proper size and shape in a board. As was the case with the two previous nut-bolt handling tests, this one was included because of its resemblance to operations which are part of shop practice.

A commonplace manipulation, packing wooden blocks in a box, and a slightly different version of this, lining blocks up three at a time on a raised strip of wood, constituted tests 20 and 21. Both were speed tests.

The next seven tests (variables 22 to 28 inclusive) were of the pencil-and-paper variety.[5] A *tracing* test required the subject to find a route over which to draw a continuous path through a maze. The maze was constructed of a series of uniform adjacent rectangles with irregularly spaced breaks in their side-walls. The penciled path had to pass through these breaks. The next variable in this group was a simple *tapping* test, for speed. A certain minimum of control was necessary here because three dots were to be made in each of a whole series of small circles. In the so-called *dotting* test much the same sort of performance was necessary. The major difference, however, was that the circles were irregularly spaced and one dot was made in each, so that more control, or eye-hand coordination, seemed to be necessary. For the *copying* test, simple line drawings were presented, and the subject's efforts to copy these were scored for accuracy in representing direction and length of the copied lines. In the center of the sheet for the *locations* test was a large square filled with rows of capital letters. Around the large square were arranged smaller ones, each containing several dots. The task of the subject was to choose, without aid of ruler or other device, the letter in the large square which corresponded to the position of the dot in the small square. In the

[5] After McQuarrie, T. W., "A Mechanical Ability Test," *J. Person Res.*, 1927, *5*, 329-337.

blocks test, various rectangular solids, apparently made of "bricks" of various sizes and shapes, were pictured for the

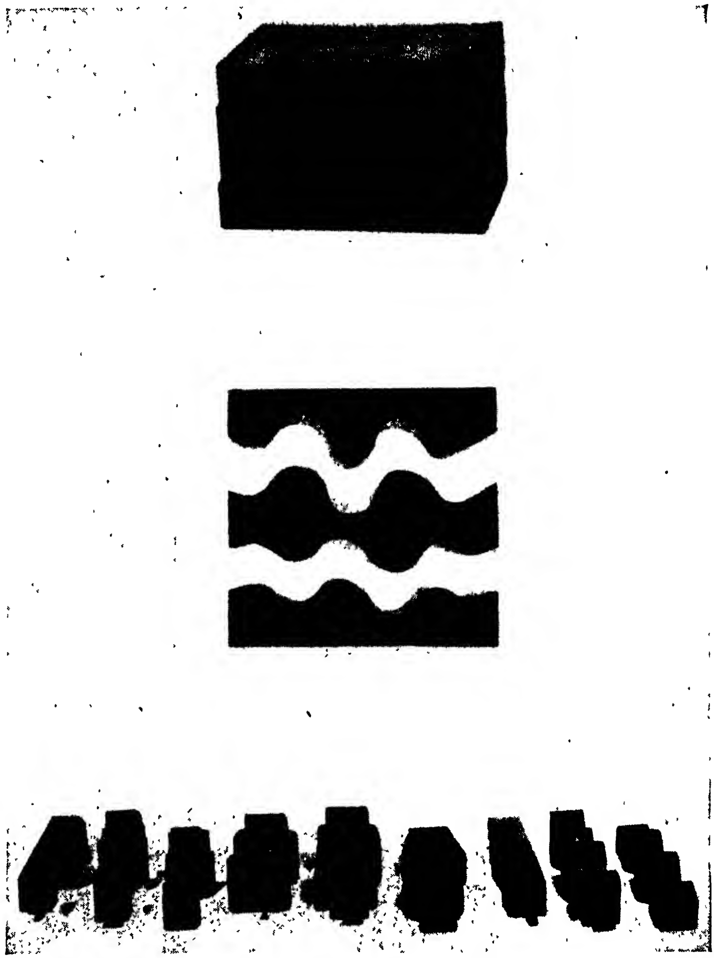


FIG. 13. Three views of the O'Connor Wiggly Block and parts. (Courtesy C. H. Stoelting Co., Chicago, manufacturers, and Williams and Wilkins, Baltimore, *op. cit.*, publishers.)

subject. His task was to report the number of surrounding "bricks" which would touch a certain one in the corresponding real solid. Finally, in the *pursuit* test, a tangled mass of lines

was shown to the subject. At the left of the figure began 10 different lines and the complicated curvings and crossings of these made up the mass in the center of the figure. At the right were shown the ends of the 10 lines. The subject had to match beginnings with ends, by visually tracing each line, and identify each with a number. These seven tests were included because they seemed to call for some of the abilities thought to be necessary for the more elaborate apparatus tests. By the use of these pencil-and-paper group tests it was easy to obtain an additional check on the factors essential in some of the tests which were administered individually.

Variable 29 was a multiple-choice picture-matching test, in which one pictured object had to be identified as belonging with, or used with, another in a series of several. In some respects this was a test of knowledge of machines and instruments.

Two of Thurstone's "lozenges" tests [6] constituted variables 30 and 31. In these pencil-and-paper tests a nearly diamond-shaped figure was shown. One of its edges was darkened and a hole through it was represented by a circle. The subject had the problem of choosing a lozenge which would be like the first one if the first were turned (in imagination) face down. The second of these tests involved a slightly modified form of the same general procedure.

Tests 4, 5, 18, and 27 were thought to demand ability to perceive spatial relations or to visualize in space. From previous work in intelligence testing the lozenges tests (30 and 31) were known to measure such an ability. They therefore serve here as a check upon the results of other tests in this general group. Test 32, not mentioned previously, seems also to belong in this group. It was called the punched-holes test, and required the subject to decide what position a hole in a sheet of paper would have if the sheet were folded in various ways.

The last three tests in the battery were taken from Thurstone's test batteries of intellectual skills. Variable 33 was word

[6] Thurstone, L. L., *Primary Mental Abilities*, University of Chicago Press, 1938, pp. 121.

opposites, 34 was word analogies, and 35 was completion. These were known from previous work to measure mainly a verbal ability, and they were included here in order to determine how they might be related to a group of tests not purporting to measure such abilities.

The final variables, 36 and 37, were ratings of the subjects by their supervisors. One was for general competence on the job, and the other was supposed to cover "mechanical ability" insofar as that existed apart from general competence.

Results. In Table 1 are shown the five factors which were necessary to account for the coefficients of correlation between all the tests.[7] The numbers in each column, as before mentioned, may be considered indices of the importance of that factor to each test. The larger the index number or weight, the more important is the specified factor in performance on a given test.

We are now ready to begin one of the major tasks in factor analysis.[8] This consists of studying the index numbers in each column in Table 1 and making a decision, or a guess if necessary, as to just what the tests with high weights have in common. In other words, to make our study meaningful, we must attempt to name the factors in mechanical ability which we have discovered by our testing and our statistical analysis.

For purposes of exposition we shall consider one factor at a time, and list for each factor the tests having highest loadings. The number of each variable in Table 1 is included here for reference purposes.

FACTOR I

(35) Completion	.561	(34) Word analogies	.480
(33) Word opposites	.547	(2) Schooling	.456

As previously mentioned, a verbal ability has been found in studies of intelligence. The completion, word opposites, and word analogies tests used in our present test battery clearly call for an ability to deal easily and readily with verbal material.

[7] This table is rearranged from Harrell, *op. cit.*, Table 8, p. 31.

[8] For a brief summary of the statistical method of factor analysis, see p. 268.

TABLE 1

FINAL FACTOR LOADINGS

	I *Verbal*	II *Youth*	III *Manual agility*	IV *Spatial*	V *Perceptual*
1. Youth	—.038	.630	.023	.068	—.012
2. Schooling	.456	.058	—.047	.079	—.059
3. Inexperience	—.177	.618	.061	.082	.033
4. Spatial relations accuracy	.282	—.043	—.194	—.178	.392
5. Spatial relations speed	.043	.058	—.018	.275	.251
6. Assembly	—.190	—.026	—.163	.383	.415
7. Form-boards	.052	.142	—.176	.517	.228
8. Interests	.149	.267	—.005	—.007	.016
9. Routine assembly	—.026	.068	—.004	.154	.624
10. Routine stripping	.001	—.090	.091	.044	.457
11. Pin-board (preferred hand)	—.036	.193	.357	—.125	.295
12. Pin-board (non-pref'd. hand)	—.185	—.160	.380	.042	—.005
13. Pin-board (with both hands)	—.187	—.013	.505	.137	—.076
14. Peg sticking	.059	.450	.270	—.063	.384
15. Nuts off	—.006	—.135	.426	.072	.123
16. Nuts on	.007	—.055	.378	—.130	.319
17. Peg sort	.110	.383	.065	—.051	.392
18. Wiggly blocks	—.050	—.073	.060	.332	.230
19. Placing bolts	—.018	—.174	.430	.058	.248
20. Block packing	—.002	.113	.384	.146	.128
21 Block strip	—.150	.200	.180	.212	—.078
22. Tracing	.094	.336	.134	.065	.167
23. Tapping	.367	.121	.322	—.001	—.015
24. Dotting	.035	.117	.501	.043	—.111
25. Copying	.109	.007	.035	.494	—.015
26. Location	.305	.129	—.059	.352	.049
27. Blocks	.077	—.008	—.062	.540	.045
28. Pursuit	.058	—.077	.077	.515	—.080
29. Picture matching	.179	.009	—.003	.222	.397
30. Lozenges A	.004	.001	.010	.327	.238
31. Lozenges B	.117	—.065	—.026	.487	—.066
32. Punched holes	.086	—.054	.034	.585	—.134
33. Word opposites	.547	—.014	.047	.234	—.044
34. Word analogies	.480	.033	.014	.271	—.048
35. Completion	.561	—.071	—.073	.244	.046
36. Poor general rating	—.257	.321	—.028	.180	—.155
37. Poor mechanical rating	—.196	.491	—.043	—.020	—.082

Since they are grouped together by their loadings here, just as in previous studies of intellectual performances, we may be fairly sure that our factor I is verbal. It would be expected that the amount of schooling the machine fixers have had would influence their verbal ability, or possibly that higher verbal ability would encourage longer school attendance, and so the fourth high loading shown above fits in with our identification. It should be noted in Table 1 that none of the supposed tests

of mechanical ability has a high loading in this factor, indicating that verbal ability as such is not important in the tests which are our major interest here.

FACTOR II

(1) Youth	.630	(14) Peg sticking	.450
(3) Inexperience	.618	(17) Peg sorting	.383
(37) Poor mechanical ability rating	.491		

Only two tests have high weights in this factor; the other high loadings are for non-test variables. From a consideration of the three highest weights we may be fairly sure that this may be called a youth factor, because youth, inexperience, and poor ratings would logically go together. That is, the younger men have less experience at mechanical trades, and, conversely, older men are likely to have better mechanical ability ratings from their supervisors. This may be because they have had more experience or because poor mechanics tend to be eliminated at the earlier ages. In other studies a maturation factor has been found, but because training more than maturation might be involved here, the factor is simply called youth. It is an interesting question as to just why the two peg-handling tests should have high loadings in the youth factor. The answer may lie in the greater willingness of the younger men to work hard at these tasks. Sorting or placing pegs might seem rather childish to the older men. Furthermore, they probably felt more secure in their positions, and did not think that the tests had anything to do with maintaining their status. But on these particular simple tests the younger men might have scoffed less, and worked harder because of the possible implications for the maintenance of status.

FACTOR III

(13) Pin-board with both hands	.505	(20) Block packing	.384
(24) Dotting	.501	(12) Pin-board, non-preferred hand	.380
(15) Nuts off	.426	(16) Nuts on	.378

Only one pencil-and-paper test, dotting, is included here and this is a manual activity. It seems fairly clear that the tests having high weights involve agility with the hands or fingers. The performances here are of a fairly routine variety. The nature of the activities required of the subject should be quite clear to him, and not much discrimination is required. Therefore his performance on the tests having high loadings here depends mainly upon how quickly and nimbly he works at the routine.

FACTOR IV

(32)	Punched holes	.585	(25)	Copying	.494
(27)	Block counting	.540	(31)	Lozenges B	.487
(7)	Paper form-boards	.517	(6)	Assembly	.383
(28)	Pursuit	.515			

Only the assembly test is non-paper-and-pencil here. It and the other tests listed clearly have one thing in common—they require the subject to think in spatial terms, or to visualize objects in space. Therefore the ability may be identified as spatial. Like the verbal factor, this ability has turned up in intelligence tests. Visualizing how parts fit together, or what a given object would be like if it were turned in space in various directions obviously have an important place in many mechanical pursuits. Unlike factor I (verbal), this ability seems logically to be necessary for mechanical as well as non-mechanical performances.

FACTOR V

(9)	Routine assembly	.624	(4)	Spatial relations accuracy	.392
(10)	Routine stripping	.457	(17)	Peg sorting	.392
(6)	Assembly	.415	(14)	Peg sticking	.384
(29)	Picture matching	.397			

This factor is the most difficult one to identify. At first glance it seems to be quite similar to factor III. There is one major difference—that the tests having high loadings here are not so routine as those listed for factor III. Since they require more discrimination, and possibly also an ability to grasp detail, this factor is tentatively called a perceptual one. Further research will be necessary to determine whether it is really not another agility factor—for there could easily be agility of various types,

so that two factors in our set of five could be of this general classification.

Summary. Five factors have been discovered in this study. Four of the factors are with some assurance identified as youth, spatial ability, agility, and verbal ability. A fifth is tentatively identified as perceptual ability. The verbal factor entered into only a small group of tests which were known measures of it; we therefore may say that it does not play a part in mechanical ability as such. The age and experience of the subjects, their ability to visualize space or think in spatial terms, their dexterity or manipulative ability, and their ability to perceive details do seem to be direct factors in mechanical performance.

Conclusion. Apparently there is no one mechanical ability. At least three and possibly four group factors in mechanical ability have been discovered in this single study. Although these four seem to account for the correlations between the variables of the present research, it is almost certain that the list will be markedly increased when other measures are used in future research. But the study here presented exemplifies a promising procedure in mapping out the field of mechanical abilities. It is a survey which charts a few of the factors, and opens the way for refinement of tests as well as further search for new or different factors. All of this adds to our understanding of mechanical abilities and at the same time tells us more about our instruments of measurement, so that we shall eventually be better equipped to tackle such practical problems as worker selection.

We should remember, however, that the factors found in this study were narrow group factors. This indicates that the total number of kinds of mechanical abilities is probably large, and that no one factor (ability) is important in very many activities. It is probable also that specific factors are important for success in each kind of mechanical performance. Therefore our analyses of mechanical ability must advance much further before we shall know completely the kind and number of abilities necessary for various practical jobs. This note of caution does not detract from the importance of our task.

CHAPTER 6

Some Factors in Conditioning of Human Responses[1]

ANYONE who tries to understand or control the behavior of another person is bound to ask, sooner or later, "How did he get that way?" Many of the answers to this question are to be found in the study of what the person has learned and how he has learned it. Since the earliest days of their science, psychologists have been formulating, testing, rejecting and accepting "laws" of learning. These laws have stated, for example, the relationship between ease of learning and type of material learned (e.g., nonsense syllables, prose, poetry). They also have dealt with the relationship between ease of learning and the manner in which the material to be learned is presented to the subject (e.g., in parts or as a whole). Still another type of law indicates the relationship between ease of learning and the characteristics of the learner (age, health, motivation, etc.). In each of these instances the psychologist seeks to state how the possible variables in the learning process may affect the success of that process. He thus has the problem of examining many different kinds of learning situations, to see what laws are important in each. He must also determine what the various kinds of learning have in common, so that general (widely applicable) statements of how learning occurs may be made. The laws which hold in a particular kind of learning are significant, then, partly for understanding that kind of learning, and partly as an aid in understanding all learning.

Some of the most significant recent advances in the study of the learning process have been made in experiments on con-

[1] This chapter is based upon the work of E. R. Hilgard and his associates. Specific papers will be cited in the appropriate places.

ditioned responses. The word "conditioned" has been used carelessly, to indicate almost any behavior learned under a wide variety of circumstances (e.g., saying that one has become conditioned to believe in the "American Way"). Psychologists, however, prefer to use the term to indicate that the learned response depends upon the particular conditions of the situation in which it develops. Thus, a constriction of the pupil of the eye would be very unlikely to come about whenever a buzzer sounds, except in a controlled experimental situation. But if a buzzer is sounded just before the presentation of a light which by itself naturally causes the constriction, the buzzer and constriction may be associated. The buzzer (conditioning stimulus) can be interpreted as signaling the advent of the light (unconditioned stimulus).[2] The constriction of the pupil is called the unconditioned response when it is elicited by the light alone. If after a good many presentations of the paired stimuli the constriction tends to occur when the buzzer alone is presented, this response is said to be conditioned.

Keeping this situation in mind, we may now give a general definition of "a" conditioned response. It is an acquired tendency to respond to a (conditioning) stimulus in a certain way because it has been repeatedly presented in a fixed relationship with another (unconditioned) stimulus which is already capable of calling out a consistent response.

The problem. The conditioned response usually has been studied as an example of learning which is of interest in its own right because it occurs in a large number of situations. There has been a trend in recent years, however, to examine the relationship between conditioning and some of the factors which are known to affect other types of learning. By the word "factors" is meant such things as the subject's knowledge of the nature of the behavior he is expected to show when

[2] There has been some confusion in these designations in the past. "Conditioning stimulus" has sometimes been synonymous with the term "unconditioned stimulus" as it is used here, and "conditioned stimulus" has been synonymous with "conditioning stimulus" as it is used here. The present terminology, however, has become standard and is used here for that reason.

various stimuli are presented to him, or attempts which may be made by the learner to prevent or enhance the behavior which the situation calls for. We should expect that these would influence the rate of conditioning, but our problem is: just how?

The more we know as to how the preparatory sets and intentions of the learner influence the formation of conditioned responses, the more we shall be capable of determining how conditioning is related to other types of learning processes.

We shall examine the following questions in turn: What effect does knowledge about stimulus relationships within the experiment have upon formation of a conditioned eyelid wink? How do instructions to inhibit or enhance the response being learned influence the acquisition of a conditioned eyelid wink?

Apparatus. In experiments based upon the protective wink reflex, one unconditioned stimulus which has been used is a slight puff of air. The unconditioned response to this is an immediate closing of the lids. The conditioning stimulus often used, one which ordinarily does not elicit a blink, is simply an easily perceived increase in the brightness of either of two "windows" (each about 10 cm. square) which are placed side by side directly in front of the subject. They are about four or five feet from him, in a room which is dark except for one other light to be mentioned later. When the window to the subject's left brightens, the puff of air follows, after an interval of .6 second. No puff follows brightening of the window on his right.

In this kind of situation the subject must acquire more than a simple conditioned eyelid wink. Since one window, the right, never is followed by a puff, it is unnecessary to wink when it changes in brightness. Therefore the subject is likely to discriminate between the two windows, making the conditioned response to one and, for economy of effort, not making (inhibiting) the response to the other. The left window may thus be called a positive stimulus, and the right window negative. A conditioned response to either window is said to occur if lid closure follows the brightening and begins before the uncondi-

tioned stimulus (puff) could have caused it.[3] Conditioned *discrimination* occurs if the subject acquires a tendency not to respond to the negative stimulus.

The details of the experimental arrangements are as follows: The subject sits with his head comfortably but firmly held in a special stand. On his upper eyelids are glued light, stiff artificial eyelashes, made of cardboard.[4] Directly before him are the two illuminated windows. At his right side is a light source which casts a beam straight *across* the front of his face at the level of his eyes. Since it comes from the side and he looks straight ahead, it does not bother him. The shadows created by the eyelashes, and the light surrounding them, fall upon a slit in a screen at the subject's left side.

Behind this screen is a pendulum. When this pendulum starts to swing in its approximately horizontal arc, it strikes various electrical contacts and thereby controls the presentation of stimuli to the subject. It also carries, in a small frame-like arrangement, a sheet of photographic paper. As this paper passes behind the opening in the screen, it is exposed to the rays from the light source at the right of the subject *except* for the areas covered by shadows from the artificial eyelashes or other markers. As the subject closes his eyes, of course, his "lashes" and the shadows they cast move downward, so that an extremely accurate and somewhat amplified photographic record of the movement of his lids is made. All this occurs during the second or two while the pendulum passes behind the opening. An additional record is made by a separate electrically controlled shadow-casting marker, to show the exact moment when the intensity of the (conditioning) light stimulus is increased. Since the pendulum moves at a known rate, the dis-

[3] The conditioned lid wink often shows a tendency to occur more and more quickly after the increase in illumination, on successive trials. This has been termed an anticipatory reaction. If the conditioned response occurs quite quickly, of course, the eye is completely closed before the puff of air can strike it, thus protecting the eye against a potentially harmful stimulus.

[4] It is surprisingly easy to get used to these artificial lashes. After a short while the subject ceases to be bothered by the fact that when he winks he moves pieces of cardboard.

tance between marks on the photographs gives the time values. The puff of air (unconditioned stimulus) is directed at the eyeball by a glass tube mounted on the head stand. The puff is released by the pendulum, which governs an electrically controlled clamp on a rubber tube leading from an air reservoir. Before each trial, pressure is built up in the reservoir to a controlled level (adjusted so as not to be painful to the subject) by an atomizer bulb. In order to permit a record as to the exact moment the puff to the eye occurs, a second tube leads from a Y-joint in the puff-tube to a marker-device which is mounted so that it too casts a controlled and movable shadow across the opening in the screen at the left of the subject.

EFFECT OF KNOWLEDGE OF STIMULUS-RELATIONSHIPS [5]

Procedure. On the first day of the experiment, training was given to develop simple conditioned eyelid responses to increases in brightness of the window which was later to be the positive conditioning stimulus (always followed by puff). During this first practice period 60 paired light and air-puff stimuli were presented; they were administered in sets of 12, with two-minute rest pauses between sets, and about half a minute between separate presentations within a set. A separate sheet of photographic paper recorded each trial.

On the second day of the experiment each subject was assigned to either a control group A or an experimental group B. Fourteen college students served in group A and 11 in group B. Records from the 10 subjects in each group whose performances (rates of learning) during the simple conditioning of the first day were most nearly comparable were chosen for final analysis. The records of the remaining five subjects were discarded, at least for purposes of this experiment.

Subjects assigned to group A received no new instructions at the beginning of the second session. The only change in the situation was that, without advance information, an occasional

[5] Hilgard, E. R., Campbell, R. K., and Sears, W. N., "Conditioned Discrimination: the Effect of Knowledge of Stimulus-Relationships," *Amer. J. Psychol.*, 1938, *51*, 498-506.

brightening of the right window occurred, never followed by the air puff. As on the previous day, the light on the left was followed by the puff. The total number of positive and negative stimuli within each set of 12 was equal; the order in which they occurred on successive trials was haphazard.

At the beginning of the second day's session, subjects in group B were told that the right window would be illuminated occasionally, but that it would not be followed by the puff, whereas brightening of the left window would be followed by the puff as on the first day. What group B thus learned from the experimenter could be learned by group A only during the course of the second sitting. How, then, does the added knowledge of group B influence their rate of learning?

After the photographs of each trial are developed, it is possible to determine whether or not conditioned responses to either stimulus occurred. The degree to which the conditioning is successful may be represented by the percentage of trials in which conditioned responses occur. This percentage may be computed for each set of 12 presentations (including six positive and six negative stimuli), so that a learning curve may be plotted within each day's trials, or an average may be computed for all 60 trials taken together.

Results. On the first day, when the subjects were simply trained with the positive stimulus, discrimination was not necessary. On the second, the *difference* between the percentage of responses to the left window (positive stimulus) and those to the right window (negative stimulus) is an indicator of discrimination between them. In Table 2 are summarized these findings in terms of (percentage) frequency of response in the control and experimental groups.

At the end of the first day's training the two groups were not very adequately matched, so that comparisons of total percentage frequencies on the second day are apt to be misleading. However, we may note that on the second day the two groups were about equal in the frequencies of response to the positive stimulus (to which they had both learned to respond on the first), but the experimental group, ***knowing*** that the

TABLE 2

AVERAGE PERCENTAGE FREQUENCY OF CONDITIONED RESPONSES TO POSITIVE AND NEGATIVE STIMULI IN SIMPLE AND DISCRIMINATORY CONDITIONING [6]

Group	*Simple conditioning, day 1, pos. stim.*	*Discriminatory conditioning, day 2* *Pos. stim.*	*Neg. stim.*	*Diff.*
A (control)	47.4	71.2	42.6	28.6
B (experimental)	57.2	70.8	29.8	41.0

negative stimulus would not be followed by puff, made fewer responses to it. In this sense the experimental group may be said to have discriminated more accurately between the two stimuli. However, the difference between the groups on day 2 is not sufficiently large to be taken as a conclusive indication of the effects of knowledge about the stimuli.

In Fig. 14 are shown the percentage frequencies of response to positive and negative stimuli during successive sets of 12 stimulations (six positive, six negative) during the two days of the experiment. The graphs for the second day are of particular interest here. These learning curves show what Table 2 could not bring out clearly: the subjects of control group A, not having had the information concerning the difference between the two conditioning stimuli, only gradually acquire a differential response. They at first respond to both the right and left windows. Group B, on the other hand, gives evidence of discriminating between the two windows from the very first trials of day 2, improving only slightly through the first four sets of 12 trials each. If we assume that discrimination is a sign of knowledge of stimulus relationships, group A may be said to have required about 24 trials to become fully aware of the distinction between the positive and negative stimuli. The responses of group B, on the other hand, show the influence of this information at once on day 2.

Summary. During the first of two experimental sessions, simple conditioned eyelid responses were established. A puff of air served as the unconditioned stimulus and the brightening

[6] Taken from Table I of Hilgard, Campbell, and Sears, *op. cit.*, p. 500.

of a weakly illuminated window as the conditioning stimulus. During the second session another window adjacent to the first

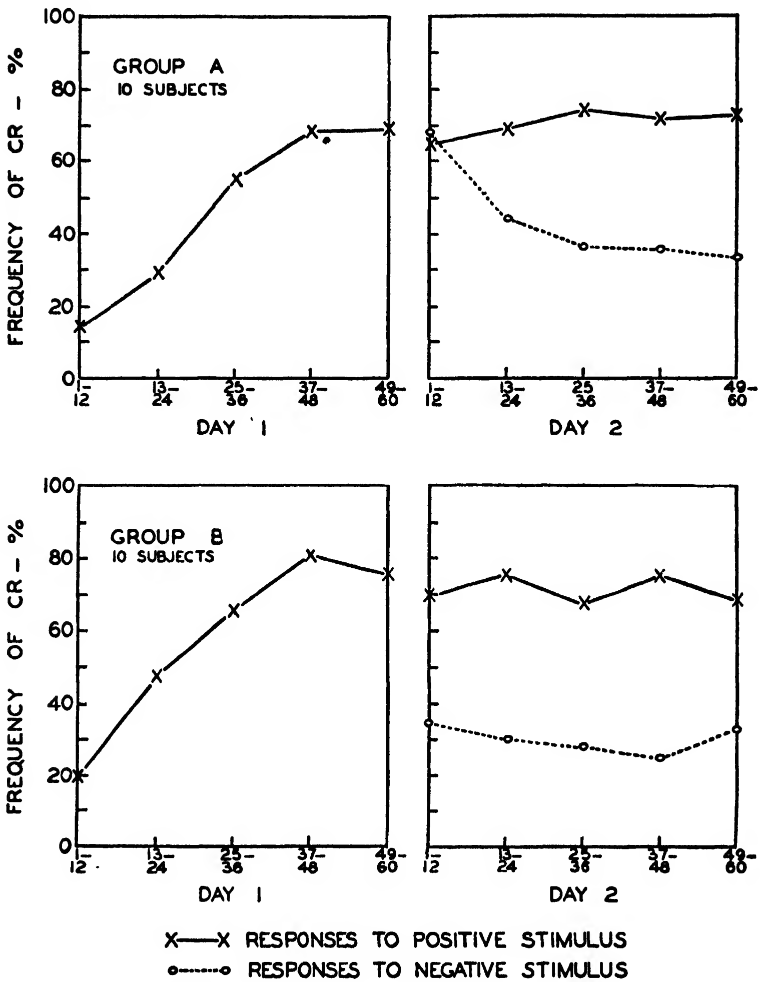


FIG. 14. Percentage frequencies of response of groups A and B to positive and negative stimuli during two days of experiment (from Hilgard, Campbell and Sears, 1938).

was occasionally illuminated also, but was never followed by the puff. One group of subjects began the second session without information concerning the purpose of the added stimulus,

or the difference between the two stimuli. Therefore they had to learn for themselves that response to one of the stimuli was unnecessary (since it was never followed by puff.) Another group were *told* that the added stimulus would not be followed by the puff.

The group without information concerning the added stimulus only gradually learned to discriminate, and to inhibit the unnecessary response; the group given advance information, discriminated between the two stimuli from the very first trials of the second sitting.

EFFECT OF SUPPORTING AND ANTAGONISTIC VOLUNTARY INSTRUCTIONS ON CONDITIONED DISCRIMINATION [7]

In the experiment just described, the subjects were simply given certain information about the experimental situation. They were in no case asked or told to do anything about it, other than to submit to the stimulation. The experiment to be reported now shows how the subject's adoption of any one of three different types of voluntary sets may influence the course of development of conditioned discrimination. The same apparatus and general procedure were used, with the changes in instructions summarized below.

Procedure. All three of the new experimental groups were given the usual first day's practice with the positive conditioning stimulus. On the next day the first of these, group C, was told about the addition of the negative stimulus, including the fact that when it occasionally occurred it would not be followed by the puff. And a significant new instruction was given: namely, to wink as rapidly as possible in response to the positive stimulus but to refrain from winking in response to the negative stimulus. Group C thus differs from group B of the previous experiment, in that group C voluntarily did something about the positive and negative stimuli which they experienced. Incidentally, it is possible that group B did this *without*

[7] Hilgard, E. R., and Humphreys, L. G., "The Effect of Supporting and Antagonistic Voluntary Instructions on Conditioned Discrimination," *J. Exper. Psychol.*, 1938, *22*, 291-304.

specific instruction. If this were the case the conditioning of group C should be about the same as that of group B.

The second group added here, D, received not only the information given to other groups, but also further instructions which were exactly the opposite of those given group C. In other words, subjects in group D were supposed to inhibit the wink response to the positive stimulus, but to wink as quickly as possible to the negative stimulus.

In both groups C and D, the tendencies to give or to inhibit conditioned responses were complicated by the fact that the subject had to remember which window was positive and which negative, and what response to make to each. In the final group, E, subjects were given the information about the nature of the stimuli, and then asked to inhibit all responses.

When the five experimental groups (including the two in our first experiment) are compared, we see that all degrees of attempts to facilitate or inhibit the acquisition of conditioned discrimination are represented. Group E is at one extreme, group C at the other, and the remaining three groups fall between these two, with group A presumably being neutral.

Results. There is a striking difference between the number of responses to the positive and to the negative stimuli in group C, as shown in Fig. 15. It is clear that the intention of subjects in group C to wink to the positive stimulus and refrain from winking to the negative resulted in facilitation of conditioned discrimination. During most of the second day, when this intention was effective, responses were made to the left window about 90 per cent of the time and responses to the right window only about 10 per cent of the time.

Group D, it will be remembered, was instructed to refrain from responding to the positive stimulus, but to wink to the negative stimulus. This meant they had to refrain from winking when the puff struck the eye, but wink when no puff occurred. Figure 15 shows unequivocally that, although this intention of the subjects produced a high percentage of response to the negative stimulus, it could *not* prevent the conditioned response to the positive stimulus. In fact, the curve of responses

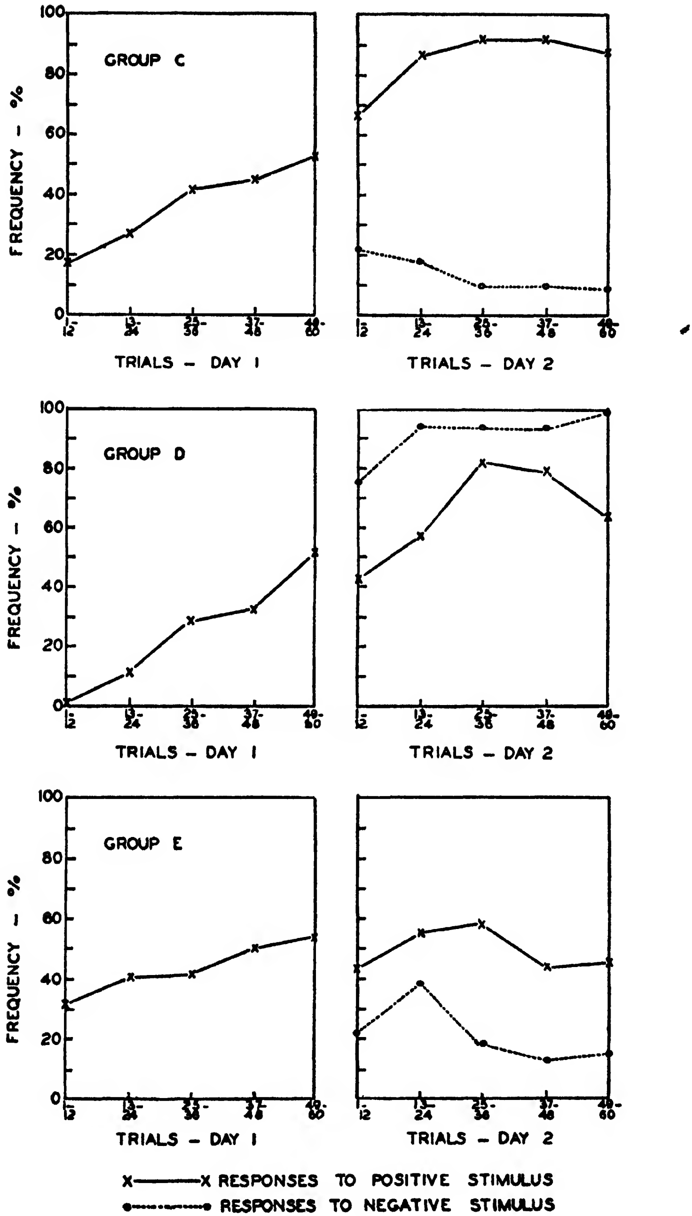


FIG. 15. Percentage frequencies of response of groups C, D and E to positive and negative stimuli during two days of experiment. (From Hilgard and Humphreys, 1938).

to the positive stimulus is at approximately the same level as the curve for group B, where no facilitation or inhibition was supposed to occur.

Something of the same sort is shown in the graph for group E. Despite their intention to avoid responding to either stimulus, these subjects hovered around the 50 per cent mark for the positive stimulus throughout the second day's session. They too showed some responses to the negative stimulus.

The three new experimental groups, taken together, show that while the intention of the subject (as determined by instructions) can have a definite influence upon acquisition of conditioned discrimination, the conditioning process is not altogether under voluntary control. In support of this statement we have the important fact that in groups D and E responses which were supposed to be inhibited occurred anyway. Furthermore, in groups C and D, where a response was supposed to occur (winking to a positive stimulus in group C, to a negative stimulus in group D), it was by no means perfect.

Summary. A conditioned discrimination between positive and negative visual stimuli was set up in three additional groups of subjects. They were given instructions to facilitate discrimination (respond to positive stimulus), to inhibit responses altogether (refrain from response to either), or to inhibit discrimination by trying to reverse the responses (respond to negative, inhibit response to positive). It was found that the intention of the subjects *did* influence the success of the conditioning process. However, the modification was definitely not great enough to permit the conclusion that the conditioning was under direct voluntary control. In fact, under certain circumstances, positive conditioned responses occur although the subject intends to prevent them.

Conclusion. It is shown here that in conditioning, as in other kinds of learning, the performance of the subject is influenced by his knowledge and his intentions concerning the experiment. On the other hand, the behavior studied here differs from certain other kinds of learning, in that voluntary sets on the part

of the subject do not *guarantee* that the conditioned response will appear if he so intends, or fail to appear if he so intends. In this sense, then, conditioned discrimination depends quite largely upon what is done to the learner, and only moderately upon his thinking about the matter.

Part Three

PHYSIOLOGICAL PSYCHOLOGY

By

DONALD B. LINDSLEY

Brown University

CHAPTER 7

Sensory and Motor Action

INTRODUCTION

During the past twenty years there has been an increasing tendency for research workers to cut across the artificial boundaries which define the limits of most scientific fields. The result in a number of instances has been the establishment of new fields of interest for research, a wider utilization of methods in use in one field by those in another, and a growing realization that there are important problems near the periphery of two adjacent fields which are often not treated adequately by either. One has only to mention the names of a few fields such as biochemistry, biophysics, neurophysiology, psychobiology, and so forth, to realize the extent to which amalgamation is taking place.

Physiological psychology, although not raised to independent status as a field, represents an attempt to bridge the gap between physiology and psychology. The subject matter in these two fields overlaps but a rather wide difference in point of view exists. In general it may be said that the psychologist is interested in the behavior of the organism as a whole, and particularly in the ways in which the behavior may be modified. The behavior may result from stimulation through a single sensory channel, or it may be the result of a complex pattern of stimuli derived internally from a multiplicity of tissue needs or externally from a busy social environment. The physiologist on the other hand is more concerned with the intrinsic functions of the various parts of the organism and their interrelationships in the maintenance of life processes such as digestion, respiration, circulation, excretion, and metabolism.

The region of greatest overlap between these two fields is the nervous system. It is through this widespread integrating net-

work that the internal and external environments influence and regulate the bodily processes in which the physiologist is interested, as they also influence the thinking, feeling, acting, and learning phenomena with which the psychologist is concerned.

Some psychologists maintain that psychology need concern itself but very little, if at all, with the physiological processes underlying behavior. They argue that physiological data are so meager and incomplete for most of the processes underlying psychological phenomena that little value is derived from a consideration of them. They prefer to describe and seek an explanation for behavior and its modifications at the behavioral level. In other words, given a stimulus which in a controlled situation will produce a certain response or modification of behavior, the essential data are the nature of the stimulus and the nature of the response. There is no concern for the intermediate steps or physiological processes by which the effects of the stimulus are translated into behavior. Such an approach is perhaps justifiable if lawful relationships can be established between all stimulus and response situations, for it might serve the essential purposes of the psychologist, namely, to be able to predict and control behavior. However, no explanation of behavior will ever be complete until the underlying mechanisms are understood. It is to this end that the efforts of physiological psychology are turned.

Of particular interest to the physiological psychologist are the processes by which receptors translate physical stimuli into nerve impulses or other means of communication within the organism, the ways in which the coordination and integration of various bodily processes is brought about when a stimulus produces a response, and finally, an understanding of the nature of, and means of identifying and measuring, the response. These factors are of importance to an understanding of the relationship between stimulus and response, and to a complete description and explanation of behavior.

The chief integrating mechanisms of the body are those of the nervous system and the circulatory system. These two sys-

tems pervade all tissues capable of response, namely, various types of muscles and glands, and provide possibilities for a widespread control of bodily activities. Thus a stimulus which causes an externally observable response in one part of the organism may also produce an unobservable, but definitely significant, response in some far-removed area.

Although the main functions of the circulatory system are those of supplying nutritive material and oxygen to body cells and the removal of waste products, more and more evidence is accumulating concerning the importance of the blood stream in the transmission of chemical agents capable of facilitating or inhibiting, and in some cases actually originating certain responses throughout the organism. The stimulating or inhibiting influences may be chemical elements normally found in the blood stream which for one reason or another fluctuate in amount or concentration and thus raise or lower the threshold of excitability of receptors, effectors, and other parts of the nervous system. In other cases the chemical agents are poured into the blood stream by endocrine glands such as the pituitary, thyroid, adrenal, and sex glands. There is also some evidence that powerful chemical substances, which may play a part in the transmission of impulses, are released in minute quantities in the neighborhood of nerve endings and synapses. For the most part, however, the chemical control of bodily responses is an indirect one since the original stimulus in most instances comes through the nervous system.

The nervous system is by far the more important means of co-ordinating bodily activities. Not only does it provide a faster means of transmitting messages from one part of the organism to another, but its lines of communication are far more extensive than those of the circulatory system. The nervous system extends to all receptors and all effectors and even controls the circulatory system by virtue of its connections with the muscular walls of the blood vessels and the heart.

In the remainder of this chapter and in the other chapters on physiological psychology some experiments will be described which throw light on the functions of the nervous system in the

transmission and integration of impulses necessary to the control of behavior. The experiments have been selected to illustrate some of the different modes of approach in the study of the nervous system. The nature and interpretation of activity in sensory and motor pathways will be described in Chs. 7 and 8. Some recent experiments in the field of audition will be presented in Ch. 9. Chapters 10 and 11 will deal with properties of the central nervous system.

ELECTRICAL PHENOMENA IN NERVES AND MUSCLES

The action current. All living tissues have electrical properties but one of the most important of these so far as the study of nerves and muscles is concerned is the development of an electric current or potential in the local region of the nerve or muscle excited to activity. This can be demonstrated easily with a nerve-muscle preparation and a sensitive galvanometer.

A suitable muscle such as the gastrocnemius in the leg of a frog may be selected and its nerve, the sciatic, carefully dissected out and cut as far centrally as possible. The nerve is then gently placed across two conducting elements or electrodes which in turn are connected to the galvanometer, a device for indicating the amount and direction of flow of an electric current in a circuit. An appropriate mechanical, chemical, thermal, or electrical stimulus applied to the nerve will set up an impulse which will pass along the nerve and cause the muscle to twitch.

As the impulse passes the nearest electrode that region of the outside of the nerve becomes electrically negative with respect to the unexcited regions ahead, including the region under the second or more distant electrode. The difference of potential between the two points on the nerve touched by the electrodes causes a current to flow in the galvanometer circuit which is registered as a deflection of the galvanometer needle as is shown in Fig. 16A. When the impulse, which is a localized wave of activity in the nerve, reaches the second electrode the nerve at that point then becomes negative to that under the first electrode which has recovered its original positive state of excitability. Thus the current flows in the opposite direction in the

galvanometer circuit and the deflection of the needle (Fig. 16C) is reversed.

Actually because the potentials in nerve fibers are so minute, often no more than a few millionths of a volt, and the frequency

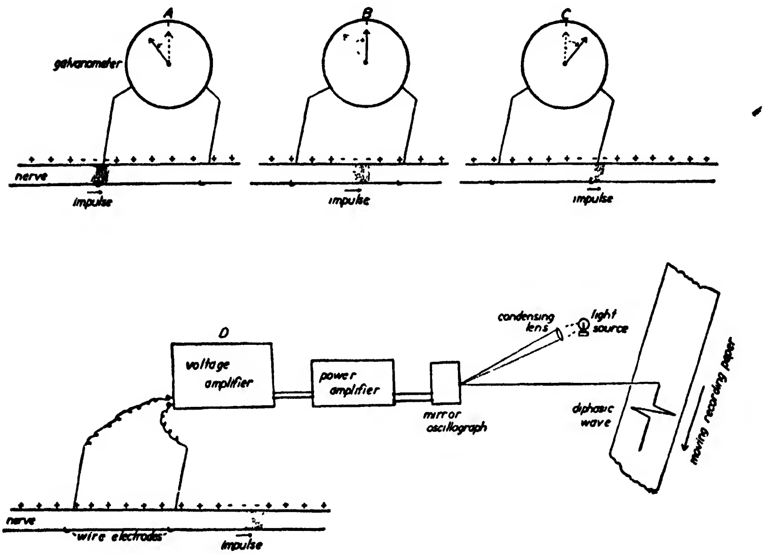


FIG. 16. Diagrams illustrating how the *action potential* associated with the transmission of a nerve impulse may be detected and recorded. *A, B,* and *C* show successive stages of the course of the impulse and the deflections of the galvanometer needle caused by the electrical variations accompanying it. *D* shows a diagrammatic arrangement of the apparatus actually used to record the action potentials. Note that as each impulse passes the electrodes a diphasic wave is traced on the moving recording paper.

of the impulses is so high, sometimes ranging up to two or three hundred or more per second, it is necessary to use a vacuum tube amplifier capable of magnifying the potentials manyfold and an oscillograph[1] capable of responding to high frequencies.

[1] Some mirror oscillographs are similar in principle to a galvanometer but are much less sensitive. The lack of sensitivity is compensated for by the use of powerful amplifiers. The most common type of instrument

A beam of light focused on the mirror of the oscillograph (Fig. 16D) and reflected to a strip of photographic paper moving at a uniform rate records the successive deflections as a two-phased wave. That is, the recording line shows a spike-like deflection in one direction when the nerve impulse passes the first electrode and a reversed wave when it passes the second electrode. Such a tracing is called an *action current* record, or more properly an action potential record since it is a measure of the potential difference between two points on the nerve when an impulse passes. A similar type of action potential record may be obtained by recording from a muscle when it is activated by a nerve impulse.

Characteristics of the nerve impulse and its action potential. The nerve impulse is a conducted disturbance of brief duration and of explosive character. It is transmitted from one segment of the nerve to another in much the same manner that the ignited portion of a fuse spreads along its pathway. Unlike the fuse, however, the nerve can be reactivated after a brief interval of time known as the *refractory period.* In the nerve the local region of activity spreads along like a wave and leaves in its wake a condition of temporary disorganization. Recovery begins almost immediately but during a portion of this state known as the *absolute refractory period* and amounting to about one thousandth of a second in mammalian nerves it is impossible to induce another impulse no matter what the intensity of the stimulus may be.

Various characteristics of a particular nerve fiber such as its size, the nature of its medullation, and its general physiological condition determine the duration of its refractory period and thereby the number of impulses which may pass over it per second. A nerve fiber with an absolute refractory period of one thousandth of a second will be able to transmit no more than one thousand impulses per second. This figure appears

used to record the detailed form of the action potential associated with the nerve impulse is the cathode ray oscillograph, a relatively inertia-less system since it involves no moving parts but only the deflection of a stream of electrons which show as a trace on a fluorescent screen.

to be about the upper limit of the largest fibers in cats or human beings and most sensory and motor pathways apparently conduct impulses at much more moderate frequencies, for the most part under one hundred per second.

Nerve impulses and their accompanying action potentials in both sensory and motor fibers are similar. The action potentials in different fibers may vary in magnitude, duration, frequency and form but they are discrete phenomena and have the same general pattern. In any one nerve fiber the action potentials are of uniform size, duration and shape providing the condition of the nerve remains satisfactory. Increasing the intensity of a stimulus applied to a receptor *does not increase the magnitude* of the action potential in a sensory nerve fiber although, within limits, it *does increase the frequency* of the impulses. This property of a single nerve fiber to give its full response or none at all is known as the *all-or-nothing principle* and applies to all individual nerve fibers. A group of fibers in a nerve trunk may, however, operate together in various degrees so as to produce, in rather small steps, wide variations in the total electrical output from the trunk as a whole.

The duration of the action potential of a nerve fiber and also the speed of conduction of its impulse are related to the size or diameter of the fiber. In a sensory or motor nerve trunk a variety of fiber sizes are found. In one of the small sensory nerves in the leg of the cat a stained cross section reveals approximately one hundred fibers varying in diameter from 1 to 18 microns.[2] The largest of these, the so-called A fibers, ranging in diameter from about 10 to 18 microns, have a conduction rate ranging from 30 to 80 meters per second. The smallest, known as C fibers and measuring only a few microns, have a conduction rate of about one meter per second. A group of B fibers of intermediate size conduct at intermediate rates. Some of these apparently belong to the autonomic nervous system.

In the larger fibers of cold-blooded animals such as the frog the rate of conduction of the nerve impulse ranges from about 15 to 30 meters per second. Some of the most recent calcula-

[2] A micron equals one thousandth of a millimeter.

tions from measurements on the ulnar nerve, studied in the arm of man, indicate that the maximum conduction rate is about 75 to 80 meters per second. Pain, which is usually considered to be conducted by the smallest and slowest fibers, presumably the C fibers, must also be carried by some of the intermediate fibers; for calculations in man, as well as measurements in other mammals, indicate that various types of pain discharges travel at rates ranging from less than one meter per second up to perhaps 15 meters per second. Touch, pressure, and temperature sensations are mediated by larger fibers than are pain sensations, generally speaking, and therefore have higher conduction rates.

Receptor function and sensory impulses. Until recent years much of our knowledge concerning the function of sensory receptors and the impulses transmitted by sensory pathways has been inferred from studies which have attempted to correlate stimulus variations with variations in sensation or with reflex responses. This meant that such investigations either had to be carried out with human subjects who could report what they felt when a particular stimulus was applied or with animals in which reflex responses could be precisely measured. In either case the information derived concerning the nature of sensory processes was indirect and inadequate since sensory impulses are subject to modification within the integrating network of the central nervous system and are somewhat removed from the experience or the reflex response.

The relatively recent application of electrical recording methods has made possible the direct investigation of the impulses in sensory pathways when various types of receptors are stimulated. The sensory nerves most accessible to study in this way are those extending from the different types of cutaneous, joint and muscle receptors (see Fig. 17) to the spinal cord. Considerable work has also been carried out with the auditory and visual pathways but because of the relative complexity and inaccessibility of the receptors and sensory pathways the results so far do not contribute as much to our under-

standing of the function of the individual receptor as do those from the simpler preparations.

Professor E. D. Adrian, a physiologist of Cambridge University, and his collaborators have made extensive studies of

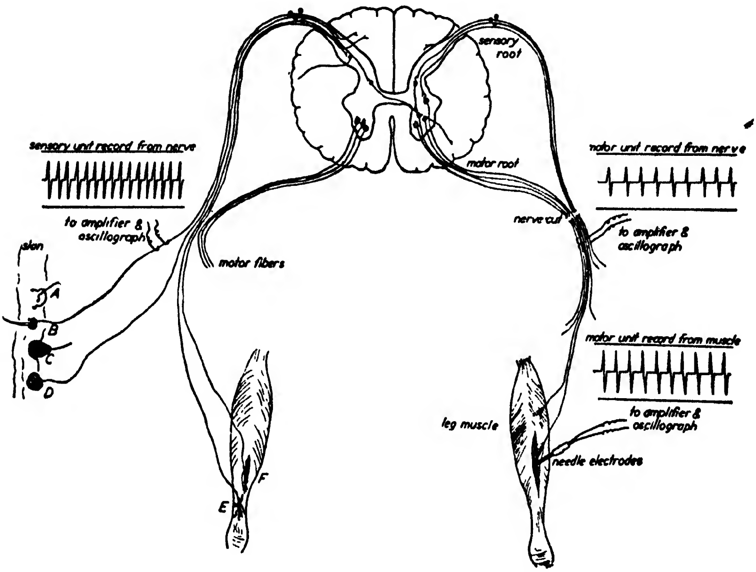


FIG. 17. Diagrammatic cross section of the spinal cord showing sensory and motor pathways. On the left, sensory pathways, receptors, and an action potential record from a single sensory nerve fiber are shown. On the right, motor pathways and two methods of recording motor unit responses are illustrated. The receptors on the left are, *A,* free nerve ending (pain), *B,* basket ending around hair follicle (touch), *C,* Pacinian corpuscle (pressure, *D,* Krause's end bulb temperature), *E,* Golgi tendon organ (tension), *F,* muscle spindle (stretch and tension). Meisner's corpuscle (touch) and a few other types of receptors are omitted.

sensory function by the electrical method. These studies not only have helped to differentiate the functions of the receptors found in the skin, tendons, and muscles but have provided direct information about the manner of operation of individual sensory units which is invaluable for the interpretation of sensory phenomena.

To cite at this point but one example, these studies have shown that sensory receptors may be grouped or classified on the basis of their rate of adaptation (decreasing activity) to a persistent stimulus of constant intensity. End-organs which respond to tactile stimulation adapt rapidly. The momentary discharge of impulses in response to a light touch on the skin lasts only about one tenth of a second. Other sensory endings which respond to pressure or pain stimuli may continue to produce impulses for several seconds. Stretch receptors, or muscle spindles found in muscles and tendons, under the influence of a constant stimulus, may continue to discharge for minutes or even hours. These facts are in general agreement with the reported sensation or the experience produced by such stimuli, but without the records of the activity in the sensory fibers there would be no way of knowing whether the momentary impression of a persistent light touch was caused by adaptation and cessation of response in the receptor or by a damping out of the effect in some region of the nervous pathway to the brain.

Since each sensory nerve trunk contains a number of fibers which carry impulses from their respective sensory end-organs it is often difficult to identify in the composite of electrical activity recorded from the trunk the action potential associated with a particular fiber and its receptor. There are however at least two solutions to this problem. Adrian and his co-workers have shown that it is possible by painstaking effort to separate and cut all of the fibers in a nerve trunk except one. It is possible to determine when the single fiber stage has been reached since but a single series of rhythmic action potentials is then obtained from the nerve trunk, central to the cut, when the whole receptor field supplied by the nerve is stimulated. In this way it is possible to study the response of a single *sensory unit,* that is, a single sensory nerve fiber and its receptor. One difficulty with the method, in addition to the time necessary to make the preparation, is that the one remaining fiber may not be attached to the particular type of receptor one wishes to study.

A simpler and perhaps better method is to seek a fine branch of a sensory nerve which supplies only one or two receptors. There are various regions among the terminal branches of sensory nerves where such a preparation may be made. Two experiments will now be described in which this method has been used to isolate and study the response of single sensory receptors. The first of these studies was carried out by Adrian and Umrath on a receptor responding to pressure stimulation. The second experiment by Matthews deals with the response of a single stretch receptor found in muscles and known as a muscle spindle.

Experiment I

THE RESPONSE OF A SINGLE PRESSURE RECEPTOR (PACINIAN CORPUSCLE) [3]

Purpose. The object of this experiment was to isolate individual sense organs of the Pacinian type and to record and study the sensory messages aroused by stimuli of different intensity and duration. In addition an attempt was made to determine whether this type of ending is responsive to both pressure and temperature stimulation as had been suggested by some writers.

Description and location of receptors. The Pacinian corpuscle (Fig. 17C) is a small oval body ranging from about 0.2 to 3 millimeters in length and is thus visible to the eye. It is a structure consisting of layers of tissue surrounding a central core into which a single nerve fiber projects. The nerve fiber loses its covering inside the corpuscle and terminates at the end of the central core in fine branches. Pressure on the capsule tends to stretch or mechanically stimulate the nerve and its fine branches.

In man these receptors are found in large numbers in the subcutaneous tissues of the fingers, the palm of the hand and the sole of the foot. They are also present in the region of ten-

[3] Adrian, E. D., and Umrath, K., "The Impulse Discharge from the Pacinian Corpuscle," *J. Physiology,* 1929-30, *68,* 139-154.

dons and joints and in the fascial tissue between muscle groups and layers. Receptors of this type are found also in and around the pancreas and in the lining of the visceral cavity.

Preparation. In this experiment it was found most convenient to work with the Pacinian corpuscles in and about the tendons in the toes of a cat. A group of several corpuscles was found consistently under a tendon in the region of the second joint of the toe. These were supplied by a fine branch nerve and were easy to stimulate by moving the joint or by pressing on the tendon. Occasional isolated corpuscles were found on the sides of the tendon.

Apparatus and procedure. The apparatus was arranged essentially as diagrammed in Fig. 16D. It consisted of a vacuum tube amplifier suitable for operation with a mirror oscillograph. Reflections of a beam of light from the oscillating mirror were recorded on photographic paper moving at a uniform rate. The records show spike-like deflections or waves, each indicating the passage of a nerve impulse in the fine branch of the nerve to which the electrodes were attached.

The cats used in the experiment were first deeply anesthetized with ether and then decerebrated; [4] thereafter anesthesia was discontinued. The tendon sheath over the second joint of the innermost toe was exposed; beneath it lay a group of 6 to 10 corpuscles. The nerve was cut centrally beyond the point where the side branch to the group of corpuscles was given off. The free end of the nerve was placed across two electrodes which led to the amplifiers and finally to the oscillograph.

Results. (1) *Stimulation of several receptors.* Flexing the toe joint or pressing on the tendon over the region of the group of corpuscles with a glass rod produced a discharge of impulses

[4] Decerebration is accomplished by inserting a probe through a hole in the skull and severing, in the region of the midbrain, the higher centers from the lower centers and the spinal cord. This permits vital functions to go on but renders the animal insensitive to pain even though the anesthesia is discontinued.

in the nerve. The action potentials accompanying the impulses were recorded. Since the nerve contained several fibers the action potentials were of high frquency, totaling approximately 250 per second. Each of the several fibers contributed to this total at the rate of 30 to 40 per second, but the discharges were not synchronous. Some of the action potentials could be distinguished by their uniform size, shape, and rhythmic sequence as the responses of a single sensory unit. Depending upon the intensity of stimulation, the individual units gave responses ranging from 5 to about 100 per second. Each change of position of the toe joint introduced a fresh discharge of impulses, rapid at first but quickly declining in frequency and ceasing to discharge entirely after a few seconds. This phenomenon is called *adaptation.* Compared to the very rapid adaptation rate of tactile (touch) receptors in the skin these receptors adapt rather slowly, but their duration of discharge is short compared to that of the muscle spindle to be described in the next experiment.

(2) *Stimulation of a single receptor.* When an isolated corpuscle on the side of the tendon was stimulated by a constant pressure with a fine glass rod, it produced rhythmic discharges at about 50 per second at the start. The frequency of discharge gradually declined to 25 per second after 5 seconds and dropped to about 12 per second at the end of 30 seconds. Initially an increase in the intensity of the stimulus applied to a single receptor increased its frequency of response.

(3) *Comparison of group and individual stimulation.* An increased pressure applied to the group of corpuscles under the tendon sheath increased both the number of active fibers in the nerve and the number of impulses in each fiber. This illustrates two ways in which a gradation of the intensity of a stimulus may affect the response in sensory pathways and presumably the strength of the sensation. One is a spatial mechanism involving the number of receptors affected, the other a temporal mechanism involving the number of impulses transmitted per unit of time.

(4) *Unresponsive to temperature stimulation.* Finally, it was found that when a metal tube in which water circulated at different temperatures, covering a wide range, was applied to the tendon over the corpuscles no response was obtained, although the same stimuli applied to the side of the toe pad, where there are temperature receptors, produced impulses in the cutaneous branches of the nerve. This indicated that the Pacinian corpuscles are responsive to pressure but not to temperature stimulation.

Experiment II

THE RESPONSE OF A SINGLE STRETCH RECEPTOR (MUSCLE SPINDLE) [5]

Purpose. The chief aims of this experiment were to record and study the response of a single proprioceptive end organ, the muscle spindle, to note the influence of the chemical environment upon its response and to investigate its adaptation process.

The receptor and its location. The muscle spindle, as its name implies, is a spindle-shaped structure. It consists of muscular, tendinous and nervous elements and is situated among bundles of muscle fibers in skeletal muscles, usually near the junction of the tendon and the muscle. Both motor and sensory nerve fibers enter it, the former ending on muscle fibers in the interior of the structure and the latter branching and forming spiral networks around the muscle fibers (see Fig. 17F). Since the elongated muscle spindles are arranged in parallel with other bundles of muscle fibers, *stretching* of a muscle puts them under tension. This stretch mechanically stimulates the fine terminal branches of the sensory nerve. A *contraction* of the muscle also mechanically stimulates the nerve endings. Thus either a lengthening of the muscle as in relaxation or stretching, or a shortening as in contraction will serve to excite the receptors.

[5] Matthews, B. H. C., "The Response of a Single End Organ," *J. Physiology,* 1931, *71,* 64-110.

Procedure. Matthews found a small muscle on the side of the middle toe of the frog which usually contains one muscle spindle, although sometimes there are two. Following each experiment the number was determined by microscopic study of sections of the muscle. A small lateral branch nerve supplies the muscle. Electrodes could readily be attached to it and cutting of the nerve was not necessary. The bone to which the end of the muscle was attached was firmly fixed by a pin stuck into the bottom of a chamber containing a fluid [a suitable physiological solution for frogs] in which the entire muscle was bathed to keep it from drying. The tendon was dissected from its attachment and a piece of silk thread tied to it. The thread extended out of the moist chamber in which the preparation was kept at constant temperature and over a pulley so that it could be weighted for applying different intensities of stretching. Thread electrodes were attached to the nerve and to wires which led to the amplifiers. The amplifying and recording system was similar to that of the previous experiment.

Results. With the application of a two gram weight which produces tension comparable to that normally developed by the muscle during contraction there appeared in the nerve a rhythmic sequence of action potentials at the rate of about 120 per second. Within two seconds the frequency of the impulse discharge declined to 60 per second and at the end of 15 seconds the frequency was about 25 per second. Thereafter the frequency declined very slowly. This is the familiar process of *adaptation.*

During the period when the receptor was discharging at the rate of about 60 per second the impulses recurred with machine-like regularity and the interval between successive impulses never varied by more than 3 per cent. However, when the impulse discharge dropped below 20 per second the impulses occurred somewhat irregularly. If after a five minute rest the two gram weight was again applied, precisely the same adaptation curve was obtained showing that the results were quite reproducible and therefore the adaptation process a stable one.

Adrian and Zotterman [6] had previously shown that for various types of cutaneous, joint and muscle receptors the frequency of discharge of impulses increases with the strength of the stimulus. In this experiment Matthews demonstrated that the frequency of the impulses set up increased with the logarithm of the stimulus, or the load applied to the muscle. This result has a bearing on the Weber-Fechner law of the relationship between stimulus intensity and sensation; namely, that a stimulus increases in a geometric progression while sensation increases according to an arithmetic progression; for it suggests that this relationship is determined by properties of the end organ rather than by some central mechanism in the brain.

By an ingenious method of re-loading Matthews was able to throw considerable light on the process of adaptation. He allowed adaptation to proceed under the influence of a stimulus to a certain point and then he removed and reapplied the stimulus. When he applied the test load at different points in the adaptation curve he found that the rate of recovery of the receptor increased more rapidly as the adaptation progressed. A further finding was that adaptation might last for several seconds after a stimulus was removed. This was called an "adaptation remainder" and was found to depend upon the intensity and duration of the stimulus.

Varying the temperature of the preparation modified the frequency of response of the receptor markedly. A rise in temperature caused an increase in the frequency of the response to a given stimulus but resulted in more rapid adaptation of the receptor. Lastly, it was found that the receptor is very sensitive to changes in the ionic concentration of the fluids around it. For example, in sodium chloride the frequency of discharge to a strong stimulus remained high, but on adding either calcium or potassium to the solution the frequency of the discharge was decreased. By adding both it was found that there was less

[6] Adrian, E. D., and Zotterman, Y., "The Impulses Produced by Sensory Nerve-endings," Part 2, *J. Physiology,* 1926, *61,* 151-171; Part 3, *ibid.,* 1926, *61,* 465-483.

diminution in frequency than by adding either one separately, thus indicating that they have antagonistic influences on the excitability of the receptor. Diminishing the calcium in the fluid environment below its normal amount increased the excitability of the receptor and prolonged its discharge of nerve impulses whereas an excess of potassium ions depressed its excitability. The same general result has been demonstrated more recently by Talaat [7] for cutaneous sensory end organs.

DISCUSSION OF EXPERIMENTS I AND II

These experiments are important because they furnish a method of studying the action and properties of a very small functional unit of the nervous system. The unit consists of a receptor which translates the physical energy of the stimulus into nerve messages and a sensory neuron which conveys them to the central nervous system where they are passed on to other units of the nervous system.

A stimulus of sufficient intensity to excite the receptor will produce an impulse in its nerve fiber. When, as in these experiments, the action potential associated with the nerve impulse is recorded it is found to have a certain amplitude, and no further increase in the intensity of the stimulus will modify its size. This is the *all-or-nothing principle* of excitation. What then is the nature of the change produced by a more intense stimulus?

In these experiments, as well as in those dealing with other types of receptors, it has been shown that an increase in the intensity of the stimulus increases the *frequency* of the discrete, all-or-nothing type of impulses and may lengthen the total duration of discharge. These facts raise two more questions: How can an increase in the frequency of the messages be appreciated or sensed as an increase in the intensity of the stimulus? Also, how can a continuous sensation of pressure or stretching be accounted for when the sensory messages are composed of successive, discrete impulses?

The answer is that somewhere in the central nervous system

[7] Talaat, M., "The Effect of Ions on the Cutaneous Sensory Endings of the Frog," *J. Physiology,* 1933, *79,* 500-507.

a fusion and *summation* of the impulses occurs. Numerous experiments have shown that summation of two kinds occurs in the nervous system, presumably at synapses or the junction of neurons. These are a *temporal summation* and a *spatial summation.* Temporal summation is the change produced at some point by the arrival of an increased number of impulses per unit of time and therefore an increased frequency of discharge produces a greater summation effect. Spatial summation, on the other hand, is brought about by the convergence of impulses from several pathways. Thus, in the experiment of Adrian and Umrath, when several Pacinian corpuscles were stimulated simultaneously by pressure upon the tendon there was a discharge of a certain frequency in each of the several nerve fibers. The convergence of the impulses in the several fibers at some central point might produce spatial summation and result in a more intense sensation than if only one of the receptors was discharging.

Experiments of the type reported here are also helpful in determining the precise nature of the stimulus which is effective for different receptors. It was shown in the first experiment that the Pacinian corpuscle responds to pressure but not to temperature stimulation. Although these receptors are activated by pressure wherever they may be located in the body it is evident that they signal very different types of information. For example, the Pacinian corpuscles located in the skin and underlying tissues of the palm of the hand or the sole of the foot signal pressure or deformation of the skin whereas those located in the neighborhood of joints, as was shown in this experiment, signal movement of the joint. Those Pacinian corpuscles in the visceral cavity presumably signal still other types of information. This raises a question concerning the adequacy of the usual classifications of receptors found in textbooks. Eventually end organs or receptors will probably be classified not only in terms of their anatomic or cell structure and the sensory messages they signal but also in terms of the characteristics of their response to stimulation, such as their threshold, frequency of discharge, and adaptation rate.

Finally, since in physiological psychology we are interested in the influence of physiological states or conditions upon behavior, the results in the second experiment of changes in the chemical or fluid environment upon the response of a receptor to stimulation should be discussed. In this experiment by Matthews [as well as in the more recent one by Talaat, previously mentioned], it is evident that a slight deficiency in the calcium content of the body fluids may serve to increase the excitability of receptors and prolong their discharge, whereas an excess of calcium has the opposite effect. An excess of potassium ions also serves to depress the excitability of the receptors. These facts emphasize the importance of maintaining a proper balance of the constituents of the blood stream. Obviously the proper equilibrium in the blood may be disturbed by a variety of factors, such as disease, faulty diet, adverse climate, and hereditary defects.

CHAPTER 8

Motor Nerve Action

ACTION POTENTIALS IN MOTOR NERVES AND MUSCLES

Just as direct information about the function of sensory receptors and the impulses in sensory nerves can be obtained by recording the action potentials in the nerve fibers, so also can an understanding of motor nerve and muscle activity be gained from the electrical phenomena recorded in these structures.

In the preceding chapter methods and experiments have been described whereby the responses of a single *sensory unit* (receptor and sensory neuron) could be recorded and studied. Methods are also available for studying the motor discharges in individual functional units of the motor or effector system, namely a *motor unit.* This is particularly important since in the case of the motor neuron the cell body and its dendrites lie inside the spinal cord and are therefore a part of the central nervous system. (In the case of the sensory unit the cell body of the neuron lies outside the spinal cord and only the axon extends into it.) Study of the activity of an individual motor unit thus not only enables one to analyze the contraction of a muscle in terms of its responding units but also affords a means of analyzing the response of a unit of the central nervous system.

THE MOTOR UNIT

A *motor unit* consists of a motor neuron and a group of muscle fibers which it controls. The cell body of the neuron is known as an *anterior horn cell* since it lies in the anterior horn of the gray matter of the spinal cord. It has short branching dendrites which form junctions or *synapses* with incoming

sensory axons and with the axons of interconnecting neurons of the central nervous system. Some of the interconnecting neurons lie wholly within the same segment of the spinal cord, some extend from adjoining segments, and some are extensions of nerve cells in various centers of the brain. A great many axon terminations from different neurons in the central nervous system impinge on the dendrites of each motor neuron.

The axon of the motor neuron transmits the outgoing impulses to the particular group of muscle fibers it controls. The number of muscle fibers that the individual neuron controls may range from a few in small muscles, such as those of the eye, to perhaps 100 to 300 in the larger muscles of the limbs. Also, in any motor nerve trunk to a muscle there are a number of fibers and each nerve fiber supplies its own group of muscle fibers. Thus for one of the larger muscles in the leg of the cat, it has been determined by microscopic examination of cross sections of the nerve and the muscle that there are about 330 *motor units* and about 150 *muscle fibers per unit,* making a total of nearly 50,000 muscle fibers. In the larger limb muscles of human subjects both the number of muscle fibers controlled by a single motor nerve fiber and the number of motor units may be still larger.

Adrian and Bronk have devised two methods for recording the electrical responses of individual motor units. One, like that for dealing with sensory unit responses, consists of cutting down a motor nerve until only one fiber remains. This method and the type of record obtained from the intact motor nerve fiber are illustrated on the right side of Fig. 17.

The other method involves recording the action potentials from a group of muscle fibers innervated by a single motor nerve fiber. It has long been known that with two electrodes attached to the surface of a muscle action potentials could be recorded whenever nerve impulses reached it and caused it to contract. The difficulty of recording in this way, however, lay in the fact that each unit or group of muscle fibers contributed action potential waves and with perhaps 100 to 200 groups active the resulting record was a confusion of irregular waves.

In order to be able to record from a single group of muscle fibers controlled by a single nerve fiber, that is, a motor unit, Adrian and Bronk devised special *concentric needle electrodes* for insertion in the muscle. These electrodes enabled them to record the responses of only those units in the immediate region of the tip of the needle (see Fig. 17). It was found that this method provided records showing discrete, rhythmic responses like those from the single motor nerve fiber. The only difference was that the potentials from the muscle were of considerably greater magnitude than those from the nerve.

Two experiments will now be described which illustrate these methods of recording motor unit responses and indicate how the motor nerves and muscles function under different conditions. The first experiment, by Adrian and Bronk, deals chiefly with motor unit responses in cats; the second, by Lindsley, illustrates the nature of motor unit responses in man.

Experiment III

THE ELECTRICAL RESPONSE OF INDIVIDUAL MOTOR UNITS [1]

Purpose. The purpose of this experiment was to investigate the motor discharge in motor nerve fibers during reflex and voluntary contractions of a muscle.

Method and procedure. The two methods just described for dealing with the responses of single motor units, namely, cutting down a motor nerve or inserting concentric needle electrodes into a muscle, were used in this experiment. Cats were used except for a few observations made in man during voluntary contraction of muscles. First the cats were anesthetized and then decerebrated. One hind limb was firmly fixed by means of drills and clamps. Several small motor nerves supplying different muscles of the limb were studied. One controls the muscle which flexes the ankle joint as in walking. Another controls the muscle of the thigh which assists in the extension

[1] Adrian, E. D., and Bronk, D. W., "The Discharge of Impulses in Motor Nerve Fibres. Part II. The Frequency of Discharge in Reflex and Voluntary Contractions," *J. Physiology*, 1929, *67*, 119-151.

of the leg as when in man it kicks out from a bended position at the knee. The nerve in each case was cut not far from its junction with the muscle and its free cut end placed across two electrodes which led to the amplifiers.

The action potentials were recorded upon photographic paper. They were also amplified sufficiently so that they could be heard as rhythmic popping sounds, not unlike machine-gun fire, from a loud speaker as each series of nerve impulses was transmitted along the nerve. The experimenters found that listening to the discharge of impulses was a distinct advantage. It enabled them to detect the stage of activity when only a single nerve fiber was discharging rhythmically.

Mechanical stimulation (pinching) of the foot set up sensory impulses which entered the spinal cord and elicited a reflex discharge in the motor nerve to a flexor muscle of the same limb. The same type of stimulation applied to the opposite foot produced a reflex discharge in a motor nerve to an extensor muscle of the original limb. Decerebration, after the condition of shock has worn off, produces a discharge of impulses in motor fibers to all extensor muscles of the hind limbs thus causing them to become rigid. This condition is known as *decerebrate rigidity.* In addition to these three methods which were used to induce reflex discharges in motor nerve fibers, a sensory nerve, on the side of the body opposite to the limb being studied, was stimulated electrically at different frequencies. This induced motor discharges in the nerves controlling extensor muscles of the limb studied.

When motor unit responses were recorded by means of concentric needle electrodes inserted in the muscles of cats and human subjects the motor nerves to the muscles were of course left intact. The electrodes consisted of an insulated wire cemented into the opening of a hypodermic needle in such a way that the potentials recorded were those which existed at any moment between the tip of the central wire and the outside of the hypodermic needle.

Results. The first thing of interest noted was that the motor discharges were very similar to the discharges in sensory

units (see diagrammed records in Fig. 17). The frequency of the impulses discharged by the motor nerve cells, like that of sensory end organs, was found to be governed by the intensity of the excitation and this in turn governed the intensity of the effect, namely, the strength of the contraction.

When stimulation provoking a flexion reflex was applied, the discharges in individual fibers to flexor muscles consisted of a series of rhythmic impulses ranging in frequency from 5 to 45 per second, depending upon the intensity of the stimulation. However, the impulses were not as regular in their rhythm as the discharges in sensory units. The action potentials from any one unit were uniform in shape and size and conformed to *all-or-nothing principles.*

The impulses recorded from the single nerve fibers to thigh extensor muscles during reflex extension or during rigidity induced by decerebration ranged from 10 to 25 per second. In these same fibers, however, a more extensive reflex discharge induced by stimulation of the opposite limb consisted of single unit impulses at frequencies up to 90 per second. Also, the number of units active increased as the intensity of the stimulation increased. Thus in reflex contractions the intensity of the contraction was found to be increased both by an increase in the frequency of impulses in individual units and by an increase in the number of units active.

The action potentials recorded from groups of muscle fibers by means of needle electrodes were similar to those found in the individual nerve fibers, except that the muscle potentials were larger. During weak or moderate contraction rhythmic and uniform responses were obtained from the individual motor units, but during strong contraction it was more difficult to keep out of the record the responses of neighboring units which tended to confuse the picture and make it difficult to follow the responses of any one unit. However, the method revealed very nicely (as in Fig. 18) how with a weak contraction the impulses first begin to appear rhythmically at about 5 per second and then increase in frequency as the intensity of the con-

traction increases. As the contraction grows stronger more and more units become active, also.

The motor unit responses recorded during voluntary con-

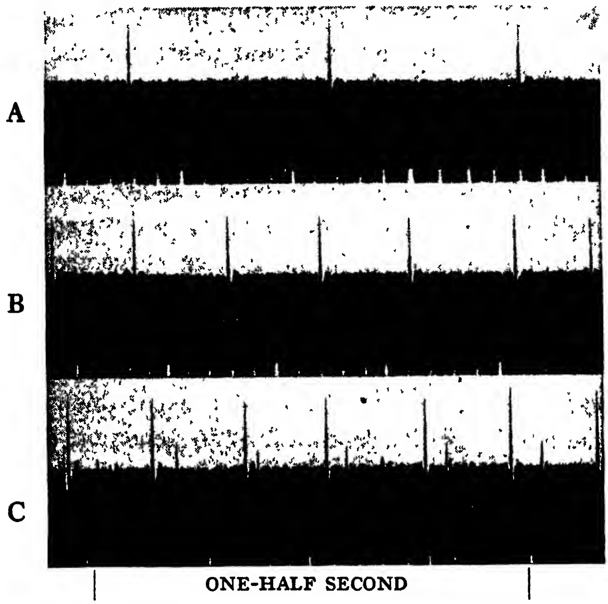


FIG. 18. Action potentials recorded from single motor units of the biceps muscle of the arm of a human subject during increasing voluntary effort. A. Weak contraction, frequency 5 to 6 per second; B. Slightly stronger contraction, frequency 9 per second; C. Still stronger contraction, frequency 11 per second, and introduction of a smaller response of a second unit at about 10 per second. (Lindsley, *Amer. J. Physiol.*, 1935, 114, p. 93.)

traction in man by means of needle electrodes inserted in the *triceps* muscle, which extends the arm at the elbow, were of the same type as those recorded during reflex contraction in the muscles of the cat. With a weak effort the frequency of the impulse discharge was low and with a very strong effort it could be made to reach about 50 per second. There was a marked increase in the number of units active throughout the muscle when a strong effort was made.

Interesting results were obtained in cats by stimulating an afferent nerve with electric shocks at various frequencies. This produced reflex motor discharges to the extensor muscles on the opposite side of the body. The frequency of the impulses of the motor discharge was increased by increasing the intensity of the stimulus but bore no particular relation to the frequency of stimulation. Electric shocks applied to the nerve at 3, 88 and 228 per second produced motor unit responses of 16, 14 and 18 per second in the individual motor nerve fibers of the opposite side. Thus somewhere within the reflex arc of a crossed reflex such as this, which contains at least two synapses, some of the impulses set up in the afferent nerve are ineffectual in exciting responses in particular motor neurons of the opposite side. In other words there is no one-to-one relation between the sensory influx of impulses and the motor outflow. However, it is probable that impulses which fail to excite one neuron may well excite another which sends a fiber to some part of the same muscle. Also in some chains of neurons it is undoubtedly necessary for incoming impulses to *summate* in order to be transmitted across the junctions or synapses. *Summation* may be either *temporal* or *spatial. Temporal summation* is due to the adding of effects of a temporal sequence or succession of impulses from the same source; *spatial summation* is due to the adding of effects from different sources.

Experiment IV

MOTOR UNIT RESPONSES DURING VOLUNTARY CONTRACTION IN MAN [2]

Purpose. The aims of this experiment were to determine (1) the range of frequency of response of which a motor unit is capable during voluntary contraction, (2) the manner in which gradation of strength of contraction occurs in various muscles and (3) the effect of sustained contractions and fatigue on the motor unit response.

[2] Lindsley, D. B., "Electrical Activity of Human Motor Units During Voluntary Contraction," *Amer. J. Physiology*, 1935, *114*, 90-99.

Subjects and muscles studied. Six normal adults served as subjects. Various representative flexor and extensor muscles in arms and legs were studied. Some of these were, the *deltoid,* which elevates and rotates the arm at the shoulder joint, the *biceps,* which flexes the arm at the elbow, the small muscle deep in the forearm which flexes the middle and ring fingers, the muscle which extends the leg at the knee and enters into the knee jerk, the muscle which extends the foot as in standing on the toes, and others.

Apparatus and procedure. The apparatus consisted of a six stage amplifier which was used to drive either a loud speaker for listening to the motor unit responses or a mirror type oscillograph for recording them on photographic paper. The arrangement of the apparatus was similar to that shown in Fig. 16D.

Four types of electrodes were used: (1) concentric needle electrodes of the type devised by Adrian and Bronk, (2) two fine insulated wires in a hypodermic needle, the potential between their bare tips being recorded, (3) a fine insulated wire with a small scratch in the insulation, which could be sewn into the muscle, and (4) two such wires with insulation removed by a knife-edge cut.

The needles were sterilized and inserted directly into the muscle to varying depths and fixed in place by an adjustable holder. No anesthesia was necessary since the chief discomfort came from piercing the skin and that was only momentary. Occasionally, however, a deep pain receptor was encountered in the muscle and required readjustment of the electrodes in order to relieve the pain.

Various mechanical arrangements of springs and weights were used for adjusting the intensity of the contraction as well as for recording it on the same record with the electrical responses.

Results. When the needle electrodes were inserted in a muscle that was not completely relaxed a rhythmic popping sound like the rattle of a machine gun was heard in the loud speaker. As the needle approached the active units the popping

sound increased in loudness and in this way one could determine when the electrodes were properly placed for recording from a particular unit. All of the subjects were able with little practice to relax completely the muscles studied so that no motor unit responses could be heard when the electrodes were inserted at different points.

From a state of relaxation with no active units in the region of the needle one is able to introduce an irregular response at a frequency of 2 or 3 per second in a single unit by making the weakest possible effort. With slightly more effort the response becomes rhythmic at about 5 per second (see Fig. 18A). Increasing the contraction slightly increases the frequency of the response to perhaps 9 or 10 per second (Fig. 18B). With a further increase in the intensity of the contraction the frequency increases likewise and the sound of other active units may be heard in the loud speaker. The new unit is usually heard as a sound of slightly different quality from the original, and the recorded response (see Fig. 18C) differs from the original in amplitude, form, or frequency. This pattern of response, namely, an increase in frequency of unit discharges and an increase in the number of units active, was found in all muscles studied when the voluntary contraction proceeded from weak to strong.

During the ordinary range of intensity of muscular contraction in the various muscles the frequency of the motor unit responses seldom exceeded 30 per second. Yet with maximal effort it was found that the frequency of the unit response reached 45 to 50 per second.[8] It is possible that in very rapid and very strong contractions the motor discharge may go even higher.

Alexander Forbes, in 1922, proposed a possible explanation of the lack of fatigue in long-sustained contractions. This hypothesis, suggesting that different units might rotate in their activity, has never been substantiated and no evidence for

[8] It was possible by means of two wires sewn into the muscle or with two wires in a hypodermic needle to isolate the response from a single unit during strong contractions.

this notion of rest awhile, work awhile, was found in the response of motor units recorded in this study. Several series of records were obtained showing that a single motor unit could be made to respond rhythmically at a frequency of 15 to 20 per second for periods lasting up to 30 minutes without any observable change in the frequency or amplitude of its response. It is somewhat astounding to realize that such a unit responding at 20 per second would in this period of time respond about 36,000 times without resting or shifting its burden to another group of muscle fibers.

However, it was found that during a strong contraction which caused fatigue within a few minutes there was evidence that the size of the action potential decreased progressively as the sense of fatigue developed, but the frequency of the response did not change and the unit did not fail to respond. The reduction in the amplitude of the response must be taken to mean that the nerve impulse which is regularly coming down the nerve is not able to excite all of the 100 or 200 muscle fibers composing the unit group. This may be due to a failure of transmission of the impulse at the neuro-muscular junction of these particular fibers (as in curare poisoning) or to some failure of the activation mechanism of these fibers. In any event it suggests that some of the muscle fibers are not contributing an electrical potential and therefore the sum of the potentials is less. Also the amount of the pull the muscle is able to exert becomes less.

DISCUSSION OF EXPERIMENTS III AND IV

Much of our behavior is reflected in the movements of the body resulting from the action of skeletal muscles. Therefore it is particularly pertinent to understand how the commands of the central nervous system are carried out by these structures. The ease, grace and efficiency with which complex movements are performed by the system of levers which compose the structural framework of man has no counterpart in all of the complicated machinery built by the hand of man.

Although much important information on motor activities

has been acquired by other methods of study, it has remained for the electrical recording techniques to reveal directly the nature of the impulse discharges which are responsible for movements of the body. From the experiments described here as well as others of similar type we learn that the impulses in motor nerves are not unlike the impulses which are transmitted from sensory receptors to the central nervous system, or for that matter are not unlike the impulses conveyed within that system.

One might inquire then how these impulses are capable of producing the smooth and delicately adjusted contractions of muscles, since each motor nerve fiber has been shown to conduct impulses of brief and discrete character. First, it has been demonstrated that the impulses discharged in motor nerve fibers during reflex and voluntary contractions occur chiefly in repetitive series of varying frequency, pattern and duration. It has been shown also that the motor behavior is made up of many units of function known as *motor units* which consist of both nerve and muscle structures.

A single nerve impulse transmitted by an individual motor fiber will produce a single twitch or contraction of the group of muscle fibers controlled by that particular fiber. However, even in the weakest sustained voluntary effort one can make the impulses are discharged in motor nerve fibers in a rhythmic sequence at about 5 per second. If impulses in all the motor fibers to a particular muscle were discharging simultaneously at this rate there would not be a smooth contraction or action of the muscle, but rather there would be a tremor-like activity in which the part moved by the muscle would jerk rhythmically in accordance with the frequency of the volleys of synchronized impulses reaching it. Fortunately for our grace of action, all of the units overlap one another in a random sequence of action which tends to make the contraction smoother.

The intensity of the contraction of each unit or group of fibers in the muscle is controlled by the frequency of the nerve impulses reaching it; an increase in frequency, within limits, causing an increase in the contraction. At a certain frequency,

which varies for different muscles, the impulses in each unit no longer cause discrete contractions with partial relaxations between impulses, but result in a fused and smooth contraction known as a *tetanus*. It should be mentioned at this point that these experiments indicate that the number of units and the frequency of discharge of each unit determines the intensity of the contraction and that variations in frequency facilitate the smoothness of the contraction. Small increases in intensity of the contraction appear to be determined by increases of frequency of discharge in each unit, but larger increases are undoubtedly mainly due to the addition of new units of activity.

SUMMARY

In this and the preceding chapter the electrical signs of the underlying activity in sensory and motor nerves and muscles have been discussed. These electrical variations are known as *action currents* or *action potentials.*

Two experiments were cited in which the function of isolated receptors (*Pacinian corpuscles* and *muscle spindles*) were studied in terms of the action potentials accompanying their sensory discharges in response to stimulation. These discharges were shown to consist of discrete impulses of rhythmic character which increased in frequency with the intensity of the stimulation and showed *adaptation* or a declining frequency with sustained stimulation. Other characteristics of the sensory discharge were demonstrated under varying conditions of stimulation.

Two other experiments were described in which action potentials were recorded from single motor nerve fibers or groups of muscle fibers during motor discharges. The nature of the impulses in *motor units* was shown to be similar to those in *sensory units*. The characteristics of the motor discharge and the arrangement of the effector mechanism in functional units were shown to account for the smoothness and gradation of muscular activity. Fatigue and other factors concerned in motor behavior were discussed in terms of the action potentials recorded from the unit structures of nerves and muscles.

CHAPTER 9

Experiments in Audition

In audition, as in most of the other sense fields, the application of electrical recording methods has contributed new information concerning the function of the sensory mechanism. To be sure, methods involving subjective judgment and report in human subjects and conditioned response techniques in animals have added much to our knowledge of auditory function. This is particularly true with respect to thresholds of acuity, limits of pitch perception and discrimination in the various species, and a host of other subjective auditory phenomena in human subjects. The conditioned response technique, as a means of testing hearing in animals, has been especially fruitful for investigating the effects of operative delimitations of the auditory system.

Under the impetus of new data revealed by electrophysiological techniques during the last ten years all methods of investigation of auditory function have been spurred on, with the result that the theories of audition have been revised and extended.

Before describing a few of the important recent experiments in the field of audition, it may be well to review briefly some of the gross aspects of the structure and function of the auditory mechanism and to state the main theories of hearing.

THE MECHANISM OF THE EAR AND ITS NERVOUS PATHWAYS

The structure of the peripheral hearing mechanism may be divided roughly into three parts, the outer, middle and inner ear, as is illustrated in Fig. 19. The outer ear in man consists of the *auricle* and the *external auditory meatus*. The auricle, or the observable part of the ear, is a structure resembling a

conch shell and it is made up of a fibrous and cartilaginous substance covered with skin. The auricle is so shaped and placed at an angle on the side of the head that it serves to collect sound waves and direct them into the curved canal

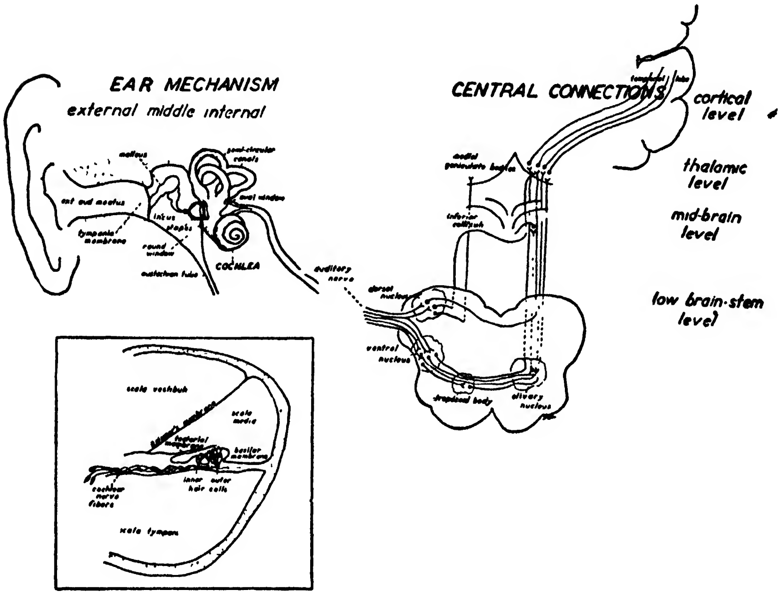


FIG. 19. Diagrammatic sketch of the peripheral mechanism of the ear and of the central auditory pathways. Inset shows a cross section of a coil of the cochlea illustrating the position of the internal and external hair cells of the organ of Corti resting upon the basilar membrane.

known as the *external auditory meatus*. About one inch from its opening at the side of the head the canal terminates in a thin, taut membrane called the *tympanic membrane*. This drum membrane, which vibrates when the pressure variations of the sound waves impinge on it, separates the outer ear and its canal from the small air-chamber known as the *tympanic cavity* or middle ear.

Extending across the middle ear cavity is a chain of small bony levers which are often referred to as the hammer, anvil

and stirrup because of their shape. This system of levers transmits the vibrations of the tympanic membrane to the membrane covering the small oval opening to the inner ear. The pressure variations thus transmitted reach the *oval window* and the fluid of the inner ear magnified about ten-fold. Two small muscles are located in the middle ear, one is attached to the wall of the cavity and the handle of the hammer in such a manner that its contraction tenses the tympanic membrane and thus reflexly protects the delicate inner ear mechanism from damage by intense sounds; the other is attached to the wall and the stirrup and presumably tenses the oval window membrane and damps its vibrations. In addition to the oval window, there is another small *round window* which opens into the inner ear. The membrane covering the round window accommodates the pressure variations applied to the fluid of the inner ear by the vibrations impressed upon the oval window membrane. At the bottom of the middle ear cavity is a duct, known as the *Eustachian tube,* which connects with the pharynx or mouth cavity and allows for an equalization of atmospheric pressure on the two sides of the tympanic membrane.

The mechanism of the inner ear is enclosed in the *cochlea,* a spiral bony cavity of about two and one half turns. The base of the cochlea in man is nearly one centimeter in diameter, but the height of the cochlea is only about five millimeters. The cone-like shell of the cochlea coils about a central bony core whose spiraling shelves resemble the threads of a screw. The *basilar membrane* and the *vestibular membrane* are attached to the edge of the shelves in such a manner as to form three divisions of the spiral tube. The two larger divisions are filled with *perilymph fluid* and join through a small opening at the apex of the cochlea. The oval window opens into one of these and the round window into the other. Between the vestibular and the basilar membranes is the third, but smaller, division of the spiral tube. It contains *endolymph fluid* and communicates with the semicircular canals of the bony vestibular labyrinth.

On the floor of this smaller division, resting on the basilar

membrane, is the *organ of Corti.* It consists of pillar-like cells which support, around the tunnel of Corti, one inner and three to four outer rows of receptor cells with hair-like projections. There are approximately 15,000 of these *inner* and *outer hair cells.* Fine branching nerves which terminate about the hair cells enter the core of the cochlea and come together at its base to form the *cochlear division of the eighth nerve.*

How the hair or receptor cells are stimulated is a disputed question, although mechanical movement or deformation is undoubtedly involved. Some writers believe that the thin jelly-like *tectorial membrane* which overhangs the inner and outer hair cells, brushes against the hairs of the receptor cells when the basilar membrane is set in vibration. Others think that the tectorial membrane has nothing to do with the process but that the movement of the fluid medium is sufficient to bend the hairs and thus stimulate the receptor cells. It is important to note that the basilar membrane is composed of striated fibers and that it increases in thickness from the base to the apex of the cochlea.

The cochlear nerve joins with the vestibular nerve to make up the *eighth cranial* or *auditory nerve.* The fibers of the cochlear division enter the brain stem, where they separate to form the *dorsal* and *ventral cochlear nuclei.* The axons of the incoming fibers from the cochlea terminate on the dendrites of new neurons in these nuclei. The axons of most of the second-order neurons cross to the opposite side near the level of entrance, although some cross over at a higher level (see diagram of auditory pathways in Fig. 19). Some fibers form synapses with motor neurons and mediate brain stem reflexes to sound stimuli. The majority of the fibers continue forward and end in the *medial geniculate body* of the *thalamus,* where they make a final junction with new neurons which carry the impulses to the *auditory area* of the *temporal lobe* in the *cerebral cortex.*

It should be pointed out that the connections made in the brain stem nuclei are very complex, as are also those at the thalamic level. There is some evidence of a well-ordered pattern of projection, point-for-point, of fibers from the cochlea

on these nuclei and subsequently on the auditory area of the cortex.

THEORIES OF AUDITION

Until some of the physiological data became available in recent years, auditory theory was based largely upon facts derived from experiments on auditory perception and what was known of the anatomical structures of the ear, cochlea and auditory pathways. Early theory depended largely upon analogy drawn between structures of the ear and physical sound producing, analyzing and resonating systems. Thus it was that Helmholtz's so-called "resonance" theory of hearing arose.

Resonance or "place" theory. This theory holds that one perceives differences in the pitch of tones because different regions of the basilar membrane vibrate sympathetically and more or less selectively to the particular frequencies of the sounds reaching the ear. Helmholtz thought that the structure of the basilar membrane, which consists of many transverse fibers arranged in parallel and increasing in length and thickness from the base of the cochlea to its apex, was analogous to the strings of a harp or to a system of resonators. He thought that each transverse fiber had a characteristic vibration frequency and would resonate whenever vibrations of that particular frequency were present in the surrounding fluid media. He also assumed that the resonance of a particular region of the basilar membrane, in response to a tone of a certain pitch, would stimulate receptor cells on the membrane which are attached to specific nerve fibers or groups of fibers. The nerve impulses set up in this way were thought to be transmitted to the brain cortex as a specific "local sign" which was identified as the pitch of the sound heard.

A more recent form of this theory holds that the intensity of the sound is appreciated in terms of the frequency of the impulses in the nerve fibers from each region of selective resonance. A further extension of the theory might also include the possibility of intensity being determined in part by the num-

ber of active fibers from any region or place of selective resonance on the basilar membrane.

Frequency theory. According to this theory the frequency of the nerve impulses arriving at the cortex determines the perceived pitch of the stimulating tone. The evidence in favor of this notion comes from certain experiments, one of which will subsequently be described, which indicate that the frequency of the action potentials recorded from the auditory nerve and central pathways, as well as in the neighborhood of the cochlea itself, increases as the frequency of the stimulating tone increases.

However, it is known that the refractory period of most somatic sensory nerve fibers is about .001 second, which, if the characteristics of conduction of the auditory nerve are not an exception, would limit the frequency of discharge in any one fiber to something less than 1000 impulses per second. This limitation, therefore, makes the frequency theory inadequate to account for the perception of tones of high pitch in terms of the frequency of impulses in an individual fiber. The solution has been sought in the possibility that it is the total frequency transmitted by a number of fibers simultaneously active but presumably carrying impulses which are not synchronized. Thus ten fibers discharging in "staggered" sequence and each transmitting ten impulses per second would give a total of 100 asynchronous impulses per second.

This raises the question of whether tones under 1000 cycles per second are transmitted by single fibers or by groups of fibers discharging at somewhat lower rates but totaling a frequency which is equal to that of the stimulating tone. Certainly for tones above 1000 cycles per second the theory must assume some sort of asynchronous temporal distribution of impulses in two or more fibers.

The theory becomes further complicated when a provision for the appreciation of different intensities of a tone is made. Since frequency is assumed to determine pitch, the only simple alternative is for intensity to be represented by the number of fibers active. However, since there is every reason to suppose

that a number of active fibers will discharge more or less asynchronously, increasing the number of active fibers would increase the total frequency of the discharge and thereby, according to the theory, would result in the perception of a higher pitch as well as a possible increase in intensity.

Resonance-Volley theory. This theory, proposed by Wever and Bray, combines elements of both the "place" and frequency theories. It assumes, on the basis of certain experimental facts, that pitch may be determined in two ways: by frequency of the nerve impulses for low tones and place of resonance for high tones. Tones of intermediate range are thought to be determined by a combination of both frequency and place of resonance.

Although this view seems at present to account for more of the experimental observations than either of the other theories, it is not necessarily, on that basis alone, the only explanation of auditory function. In fact, there are several other possible ways of explaining the same experimental data in terms of frequency, place of resonance, and number and pattern of active fibers. Although one may choose to follow the law of parsimony, or the simplest explanation of the observed facts, it is nevertheless true that the histological pattern of the inner ear and the projection of its nerve fibers on the cochlear nuclei is extremely complex. Thus it may be that our final conception of the manner of operation of the auditory mechanism will not be a simple one.

DESCRIPTION OF EXPERIMENTS

Experimental studies of auditory function during the past decade have been so numerous and have approached the problem in so many different ways that it is difficult to select a few representative experiments. The importance of any particular study cannot be completely evaluated at present and undoubtedly must depend upon the point of view from which it is considered; that is, its contribution to technique, peripheral or central analysis, theory, and so forth. It seems certain, however, that greater progress has been made in experimentation

in the field of audition during the past ten years than during any comparable previous period.

In this chapter four experiments will be briefly sketched. These represent, for the most part, either different techniques or different areas of investigation. Each of the studies contributes some important fact or series of facts which further our understanding of auditory function.

The first experiment to be described was carried out by Wever and Bray of Princeton University. They demonstrated that the electrical responses recorded from the auditory nerve of a cat, when amplified and turned back into sound, reproduce the tone or speech introduced into the cat's ear as a stimulus. This investigation was important not only because it was a well-planned and executed study which provided some new facts of importance to auditory theory, but because it opened up a whole new realm of problems in the electrophysiology of hearing and stimulated workers to renewed interest and activity in the field of audition.

The results of this study were subsequently confirmed, clarified and extended in other laboratories, notably by Adrian and associates in Cambridge, England, Davis and collaborators in the Harvard Medical School and also by Wever and Bray and their students, who have continued a systematic program of investigation in this field.

The second experiment to be described deals with the problem of "stimulation deafness" and its correlations with functional and anatomical changes. This study represents a collaborative effort by several workers in different laboratories. Davis and Derbyshire of the Harvard Medical School were largely responsible for the electrophysiological tests of hearing loss in the animals which had been exposed to loud tones for long periods of time. The exposure of the animals to prolonged stimulation and the application of certain conditioning tests was done by Kemp, then of Clark University, and Upton of Harvard University. Lurie of the Harvard Medical School made the final histological studies of the damage produced in the cochlea of the guinea-pigs.

Stevens, Davis and Lurie, in the third experiment, recorded the electrical potentials generated at various points on the cochlea when tones of different frequencies were used as stimuli. They also studied the effect of operative damage to different regions of the basilar membrane. This experiment throws light on the localization of pitch perception on the basilar membrane.

The final experiment, by Ades, Mettler and Culler is one of several studies on audition carried out by Culler and his associates, first at the University of Illinois and later at the University of Rochester. This study deals with the central pathways over which impulses originating in the cochlea must pass on their way to the cortex of the brain.

EXPERIMENT I

ACTION POTENTIALS FROM THE AUDITORY NERVE OF THE CAT [1]

Purpose. The purpose of this experiment was to attempt to lead off action potentials from the auditory nerve of the cat during stimulation with tones of various frequencies and to determine whether there was any relationship between the frequency of the action potentials and the frequency of the stimulating tone.

Operative procedure. Because the auditory nerve lies deep in the neck tissues at the base of the skull and is only about 2 to 6 millimeters in length from its point of exit in the cochlea to its entrance into the medulla, the operative procedure for exposing it constituted one of the major problems in this experiment.

When the cat had been placed under deep ether anesthesia, the animal was decerebrated by inserting a blunt probe through a hole made in the skull and severing the brain system. After decerebration the animal is insensitive to further operative

[1] Wever, E. G., and Bray, C. W., "The Nature of Acoustic Response: The Relation Between Sound Frequency and Frequency of Impulses in the Auditory Nerve," *J. Exp. Psychol.*, 1930, *13*, 373-387.

procedures and the anesthesia, which might have depressed activity in the auditory pathways, was discontinued.

From the opening in the skull made for the decerebration, the bone and tissue at the junction of the skull and neck was further resected back over the lateral surface of the cerebellum. Part of the cerebellum was carefully removed, exposing the auditory nerve. A small copper wire hook, which was to serve as the "active" electrode, was then gently placed under the auditory nerve. The "inactive" or "indifferent" electrode (so-called because it is placed on inactive tissue) which completed the circuit through the animal was attached either to the severed cerebral tissue or to the muscles of the neck.

Apparatus. The action potentials from the nerve were conducted by means of shielded cables to amplifiers which were housed in a sound-proof room sixty feet away. Here they were amplified from a few hundred to as much as 100,000 times, although usually amplification of about 6000 fold was sufficient. The amplified signals were then listened to by an observer through telephone receivers or viewed on the screen of a cathode ray oscillograph.

Stimuli. In order to avoid possible induction effects, electrical sound generating devices such as an oscillator were avoided, and wind sources of sound such as the Galton whistle, organ pipes and the human voice employed to produce the sounds which were led into the cat's ear. Sound frequencies, ranging from 105 to 5200 cycles per second, were employed.

Results. The foremost fact discovered was that when a tone of a certain frequency was introduced into the cat's ear, the frequency of the electrical changes or action potentials from the auditory nerve corresponded well enough for the observer to identify it. This held for the entire range of sound frequencies used. Even words spoken into the cat's ear could be recognized and the speaker identified. This meant that the physiological mechanism represented by the cat's auditory receptor system and the auditory nerve were reproducing the electrical counterpart of sounds in a manner similar to that of a transmitter in a sound communication system.

In addition to the corresponding increase in frequency of the electrical responses with an increase in the frequency of the stimulating sound, some rough tests of the transmission of intensity changes in the stimulating tone were made. These showed that within limits the louder the stimulus the louder the signal heard by the observer. Thus there was indication of at least a gross correlation between the degree of activity initiated in the nerve trunk and the intensity of the stimulating sound.

Finally a rather laborious series of detailed control experiments were undertaken to prove that this phenomenon of transmission of sounds by the cat's auditory system was really physiological rather than due possibly to some kind of artifact. *Induction effects* were ruled out by using non-electrical and non-magnetic sources of sound and by the use of adequate shielding of the wires between the animal and the amplifiers. *Microphonic effects* caused by vibration of amplifier tubes by the stimulating sounds was prevented by housing the amplifiers in a soundproof room at a distance. *Mechanical vibration of electrodes* as a possible source of the response was ruled out by placing the electrodes on tissue not in the immediate vicinity of the auditory nerve and failing to obtain the response.

It is notable, however, that the response was also obtained from skin, bone or muscle near the nerve and from the cerebellum and the region of the brain stem near the entrance of the auditory nerve into the medulla. This was interpreted as spread of the electrical potentials generated in the nerve. The response was in every case abolished by death of the animal, or by the destruction of both cochleas; it was also abolished or reduced by restriction of the blood supply to the head. These facts were interpreted as indicating the physiological nature of the response. Still other experiments were devised to test the possibility that mechanical vibrations from the cochlea might be transmitted along the auditory nerve.

Discussion. Tests were made by Wever and Bray which seemed to them to indicate that the source of the electrical responses they were recording were associated with the con-

duction of impulses in the auditory nerve and were not due merely to the spread of electrical effects from the receptor mechanism in the cochlea. Later, Saul and Davis, and Derbyshire and Davis demonstrated that the electrical activity recorded from the auditory nerve with unshielded electrodes is apt to contain a mixture of the nerve response and the cochlear response. For the most part, this finding probably does not disturb the essential significance of the Wever-Bray experiment, namely, that within limits, increasing the frequency of the stimulating sound increases the frequency of the action currents which may be recorded from the auditory nerve trunk.

A crucial, but difficult, experiment still remains to be done in order to demonstrate the relationship between frequency of a stimulating tone and the frequency of action potentials in a single fiber of the auditory nerve. This experiment is faced with several obstacles. The position and shortness of the nerve almost preclude the application of techniques for cutting it down. Also any such attempt is apt to interfere with the cochlear artery which runs along the nerve. Attempts to insert a fine, insulated wire electrode into the nerve have not been successful in isolating the responses of a single fiber.

Experiment II

INVESTIGATIONS OF "STIMULATION DEAFNESS" [2]

Purpose. This series of experiments was planned to answer the following questions: Do animals exposed to loud tones for long periods of time show a loss or reduction of hearing as measured by recording the electrical responses of the cochlea and by conditioned reflex tests? If so, does the frequency at which hearing loss occurs correspond to the frequency of the exposure tone? Does histological study of the cochlea reveal damage to a local region of the organ of Corti on the basilar

[2] Davis, H., Derbyshire, A. J., Kemp, E. H., Lurie, M. H., and Upton, M., "Functional and Histological Changes in the Cochlea of the Guinea-Pig Resulting from Prolonged Stimulation," *J. General Psychol.*, 1935, *12*, 251-278.

membrane which might be correlated with the functional loss and therefore support a "place" theory of hearing?

Method of exposure. Approximately fifty guinea-pigs and a few cats were subjected to intense tonal stimulation for 15 to 24 hours a day for periods of time ranging from 10 to 75 days. The frequencies of exposure tones used in different series of experiments were 600, 800, 2400 and 2500 cycles per second. The intensities of the sounds ranged from 65 to 106 decibels (a logarithmic unit of sound intensity which is roughly equivalent to a "sensation unit" or a just-noticeable-difference) above the human threshold. These intensities range roughly from the equivalent of the sound energy in the continuous din of traffic on a busy street to that produced by riveting hammers in a boiler factory. The tones were usually generated by oscillators and amplified so as to activate a loud speaker placed directly over the animals in a small soundproof cage. In a few experiments an Edelmann whistle activated by a pressure pump was used to produce a tone of 2400 cycles per second.

Electrical testing of hearing. Once the animals had undergone a period of prolonged stimulation they were operated upon so that either the auditory nerve or the promontory of the round window were exposed. Although the recording of action potentials from the auditory nerve was attempted in a few experiments by inserting needle electrodes into the nerve, most of the testing was done by recording the electrical responses from the region of the round window. The "active" lead was a wick electrode touching the promontory of the round window; the "indifferent" lead or electrode was attached to the neck musculature. Leads from the round window give the cochlear response rather than the response of the auditory nerve.

Beat frequency and dynatron oscillators were used to generate tones for testing frequency levels between 60 and 10,000 cycles per second. These tones were amplified and led into one of two loud speakers capable of responding faithfully to this range of frequencies. The sound from the loud speakers was introduced into the cat's ear from a side arm of a considerable

length of coiled rubber tubing, which prevented the development of resonance effects and harmonics in the stimulating tone.

The electrical potentials led from the auditory nerve or the promontory of the round window were amplified and reproduced as standing waves on the luminous screen of a cathode ray oscillograph tube where they could be photographed or observed and measured directly on the face of the screen. The intensity of the tone led into the cat's ear which produced a just-visible wave (about one microvolt) on the screen of the oscillograph was assumed to represent threshold intensity for that particular frequency. Thresholds were thus determined for representative frequencies throughout the test range. These were plotted as deviations from the average normal human curve of hearing sensitivity determined by the same system. In determining the degree of hearing loss, the "audiograms" (audibility curves) for the experimental animals were compared with those of a control group not previously exposed to intense tones.

Finally when each experimental period of testing was completed the animals were sacrificed, the ear mechanism was removed, sectioned, stained and examined histologically.

Results. In three series of animals exposed to tones of 95 decibels or below for varying periods of time there was either no loss or very little reduction of hearing as measured by the electrical testing method or by a conditioned reflex testing method employed by Kemp, although in one or two of the animals the conditioned reflex method showed a significant loss. In all cases of slight hearing loss the reduction of sensitivity extended over a rather wide range of frequencies. In only one of the animals was there evidence of damage to the cochlear mechanism upon histological examination. In this animal some external hair cells showed degeneration in the second, third and fourth coils of the cochlea.[3]

In the fourth and fifth series of animals, exposed to tones of

[3] The cochlea of the guinea-pig contains a greater number of coils than does the human cochlea.

2400 or 2500 c.p.s.[4] at 95 decibels or more, practically all showed a significant and marked hearing loss. In most of these animals there was a widespread or generalized loss of acuity for tones covering almost the entire range of test frequencies, although the greatest degree of impairment was in the neighborhood of 1200 cycles per second.

Histological study of the cochleas of these animals revealed damage, in the more moderate cases, consisting of degeneration of external hair cells in the middle of the second cochlear whorl and in a few of the more extreme cases, of widespread degeneration of external hair cells, hemorrhage and breaking away of the vestibular membrane and the inner rows of external hair cells.

The authors conclude that these experiments confirm the findings of Wittmaack, Yoshii and others that prolonged exposure to intense tones will produce damage to the organ of Corti, but that tones above 95 decibels are usually needed and that tones of 2500 c.p.s. are more effective than lower tones. In general their results agree with the findings of Horton, who utilized the conditioned reflex method of testing hearing and showed that loss due to prolonged stimulation tends to be spread over a wide range of frequencies.

The fact that the most consistent functional losses centered around 1200 c.p.s. (although the exposure tone was 2500 c.p.s.) and the most consistent and extensive damage occurred in the middle of the second coil of the cochlea led them to conclude that there is localization of "pitch perception" on the basilar membrane, that is, a localization in terms of a "zone" rather than a specific "place." According to the authors "the results indicate that the electrical response of the cochlea is generated by the hair cells and also support the 'place' or resonance theory of cochlear function."

[4] C.p.s. stands for cycles per second.

Experiment III

THE LOCALIZATION OF PITCH PERCEPTION ON THE BASILAR MEMBRANE [5]

Purpose. The object of this experiment was to determine the spatial representation (region of sympathetic resonance) on the basilar membrane for a wide range of audible frequencies.

Method and procedure. Guinea-pigs were used as the experimental animals. By way of method, it was first necessary to demonstrate that the threshold of recorded electrical response from the cochlea (a physiological index) may be taken as representative of the threshold of actual hearing or perception in the animal. This was inferred from several lines of evidence, two of which are: the similarity of threshold curves of hearing in man determined by reported sensation and those determined for the guinea-pig by minimal electric response from the cochlea; and the essential correspondence between threshold curves for the guinea-pig determined by the electrical method and by conditioned reflex methods.

After exposing the cochlea through an opening in the bony case which partially surrounds it, an "active" electrode of the wick type was placed in contact with some part of the cochlea. The "indifferent," or grounded, electrode was attached to the neck muscle to complete the circuit through the animal. The electrical potentials from the cochlea were led off to an amplifier and a cathode ray oscillograph where the magnitudes of the potentials were measured. The method of generating stimulating tones of different frequencies and leading them into the ear of the guinea-pig was the same as in the preceding experiment.

Except for some preliminary observations, the general procedure of the experiment was to determine the threshold of the

[5] Stevens, S. S., Davis, H., and Lurie, M. H., "The Localization of Pitch Perception on the Basilar Membrane," *J. General Psychol.*, 1935, *13*. 297-315.

electrical response from the cochlea for 26 tones between 60 and 12,000 cycles per second. The next step was to damage the cochlea locally by drilling a small hole through its wall at some point along its length and then to redetermine the threshold of response for all stimulating tones. Later the cochlea was fixed, sectioned, stained and examined microscopically to determine the precise region of damage. The difference between the normal thresholds and those after operative damage to the cochlea were then correlated with the exact region of injury along the basilar membrane.

Results. Preliminary observations in a few animals showed that there was some degree of localization of pitch along the basilar membrane since the threshold for high tones was lower at the basal end than at the apex; the reverse was true for low tones. Potentials generated in the cochlea were found to leak out most readily through the round window opening at the basal end of the scala tympani and through the thin bone at the apex of the cochlea. Potentials from the cochlea appeared to be generated in the hair cells of the organ of Corti.

In twenty guinea-pigs tested by the methods outlined above, both before and after damaging the organ of Corti, good correlations were found between the regions of the injury on the basilar membrane and the changes in threshold. By plotting the frequency range within which deviation from normal sensitivity occurred in each guinea-pig against the particular area of the basilar membrane in which injury was recognized by microscopic study, a curve was obtained which shows that the octaves at different regions of the audible scale of tonal frequencies do not correspond to equal segments along the basilar membrane. The lower octaves are much more closely grouped than the higher octaves. High tones were found to be localized at the basal end of the cochlea and low tones at the apical end with tones of about 2000 c.p.s. localized near the middle.

By integrating and plotting some data by Shower and Biddulph on human pitch discrimination so as to extend it throughout the audible range the authors arrived at the conclusion that

the human ear can distinguish approximately 1300 tones. Since there was a very close correspondence between this curve and the one showing localization of pitch along the basilar membrane for the guinea-pig the authors further concluded that a discriminable pitch difference is roughly equivalent to a distance of .025 millimeters on the basilar membrane or the space occupied by two external hair cells. Any attempt to relate the number of external hair cells in a single longitudinal row along the basilar membrane to the number of just-noticeable-differences (j.n.d.'s) in pitch is necessarily complicated by the manner of innervation of the hair cells, since each nerve fiber supplies one or more cells.

In one animal in which the internal hair cells were intact but the external hair cells degenerated there was a 30 to 40 decibel hearing loss for all tones. This was interpreted as indicating that the internal hair cells have a higher threshold than the external hair cells and that they are less vulnerable to damage by loud sounds. If it is true that the internal hair cells have a threshold which is 30 to 40 decibels higher than the external hair cells, it must mean that the internal hair cells are seldom stimulated by the ordinary range of sound intensities. Consistent with this is the fact that they are so placed on the basilar membrane as to be affected only by its most extensive vibrations.

Experiment IV

THE EFFECT OF CENTRAL LESIONS ON HEARING IN THE CAT [6]

Purpose. The aim of this experiment was to determine the effect of lesions (tissue destruction) in the medial geniculate bodies of the cat's brain on pitch perception as measured by the conditioned reflex method.

Method and procedure. First, young cats were conditioned to respond by movement in a rotating cage to several frequencies, at octave intervals, between 125 and 8000 c.p.s. The un-

[6] Ades, H. W., Mettler, F. A., and Culler, E. A., "Effect of Lesions in the Medial Geniculate Bodies upon Hearing in the Cat," *Amer. J. Physiol.*, 1939, *125*, 15-23.

conditioned stimulus (shock) was avoided by the animals when they learned to respond, by moving, to the preceding conditioned stimulus (tone). Testing of the conditioned responses proceeded in this way for several days until consistent limens or thresholds were obtained. The animals were then operated upon under anesthesia. An opening was made in the skull directly over the geniculate bodies and an electrode (sharp tungsten wire insulated except for the very tip) was inserted for making electrolytic lesions in the medial geniculate bodies. An "inactive" electrode was attached to the scalp musculature to complete the circuit. The tip of the "active" electrode was placed in the medial geniculate body in a localized position, by means of a Horsley-Clarke stereotaxic instrument. This is a device which fits over the head of the cat in such a way that an electrode may be inserted into a submerged region of the brain at a precise point previously computed from measurements.

With the electrodes in place, a three-milliampere current was turned on in the circuit for 20 to 30 seconds. This focused the current at the desired point at the tip of the "active" electrode and burned the immediate tissue of the geniculate body, making what is known as an electrolytic lesion. Later, after post-operative testing of thresholds for the various stimulus tones by the conditioned reflex method, the cat was killed and histological sections were made of the medial geniculate bodies in order to determine the precise location and extent of the lesions.

Results. Twelve cats were studied. In the first two the lesions were only unilateral, affecting only one of the medial geniculate bodies. In the ten remaining animals the lesions turned out to be bilateral and for the most part symmetrically placed. Retesting of thresholds of the two animals with unilateral lesions revealed no significant differences from preoperative tests. Since the auditory pathways on one side remained intact, this result was to be expected. In the ten animals, each with bilateral and symmetrically placed lesions,

there was a significant elevation of the threshold of hearing for one or more of the tones used as test stimuli.

The "audiograms" for the ten animals showed some loss of auditory acuity for all test frequencies, but there was usually one region of the test range which showed a particularly marked loss. In general there was progressively less loss for the tones on either side of this particular frequency. The important point to note is that there was a definite correlation between the frequency with the most marked elevation of threshold and the specific region of the medial geniculate body in which the lesion was found in each case. The specific loci in the medial geniculate bodies for the various frequencies was as follows: 8000 cycles, dorsal region; 4000 cycles, anterior region; 2000 cycles, lateral region; 1000 cycles, posterior region; 500 cycles, medial region; and the lower frequencies on the ventral surface.

The investigators interpret these results as indicating that there are specific pathways leading into the medial geniculate bodies which mediate particular frequencies of pitch perception. They further believe that there is an orderly projection of the fibers from the organ of Corti in the cochlea upon the medial geniculate bodies. They cite other evidence which suggests that this point-for-point projection of the areas of the organ of Corti probably extends to the auditory area of the cerebral cortex. Finally, they see in these results evidence of a specific routing by pathways of impulses concerned with the perception of pitch, a fact which suggests to them that pitch discrimination must occur at lower levels, presumably in the cochlea.

SUMMARY

Four experiments have been described from the recent literature in the field of audition. These are only a few of the many experiments which, during the past ten years, have contributed a wealth of new information about the working of the important sensory mechanism of hearing. The first experiment in this group, by Wever and Bray, which was largely responsible for inspiring much of the subsequent work in the field,

made a discovery which at first appeared to reduce the problem of pitch discrimination almost entirely to a basis in line with the *frequency theory* of hearing. Later, these same experimenters, in the light of further investigation, proposed a *resonance-volley theory* which combined elements of the frequency and resonance or "place" theories.

The second and third experiments, as well as a number of others not described, bring out evidence of the localization of pitch perception on the basilar membrane, which strongly reinforces the *resonance* or *"place" theory*. But even the investigators involved in these studies have found evidence that within certain limits, in the lower range of frequencies, the number of impulses per unit time (frequency) is responsible for pitch perception. Opinion of various workers in the field seems to be in general agreement at present that some form of frequency-resonance theory is demanded to explain all of the known facts of audition. The precise nature of the details of such a theory have not as yet been agreed upon.

The fourth experiment is important for introducing new evidence of an orderly projection of the peripheral receptor system in the inner ear on a nucleus (the medial geniculate body) of the central auditory pathways, and for demonstrating that pitch perception even centrally is confined to certain definite channels or pathways, an implication which to these investigators is further proof of the peripheral discrimination of pitch.

CHAPTER 10

Cortical Functions

In the preceding chapters of this section some electrophysiological experiments were described which show how impulses originating in sensory receptors are transmitted to the central nervous system and how impulses emanating from the central nervous system evoke responses in effector organs. The transition from sensory messages to motor impulses is accomplished by adjustments in various parts of the central nervous system. This process of integration begins at the level of the spinal cord, where relatively simple reflex adjustments are made, and extends to the higher centers of the brain where adjustments responsible for the more complicated aspects of behavior occur.

Except for the spinal cord and the ganglia of the autonomic nervous system, the integrating centers of the nervous system are concentrated in the brain and are confined to the cranial cavity, although their spheres of influence extend to all parts of the organism. Of the numerous centers in the brain, the cerebral cortex exercises the highest degree of integration and maintains a certain measure of control over all of the lower centers. From the point of view of phylogenetic development of the nervous system the cerebral cortex is the most recently acquired structure and is found only in vertebrates, being minimally represented in certain fish and attaining its highest development in man. In its ontogenetic or embryologic development in man the brain recapitulates, essentially, the stages of its development phylogenetically.

SOME ASPECTS OF BRAIN STRUCTURE AND FUNCTION

The brain, like the spinal cord, is composed of gray matter and white matter. White matter consists of bundles of nerve

fibers which transmit impulses from one part of the nervous system to another. Gray matter mainly consists of nerve cells, their junctions or synapses with other nerve cells, and supporting structures. Aggregations or groups of associated nerve cells are referred to as *nuclei* or *ganglia,* and are usually concerned with the control of specific functions.

The interior of the spinal cord is composed of gray matter, consisting of motor cell bodies (sensory cell bodies lie in ganglia just outside the spinal cord) and interconnecting neurons, and the outer portions of it are composed of white matter or conducting pathways. This relationship is reversed at the level of the higher brain centers. There a convoluted or infolding mantle of gray matter, the *cerebral cortex,* envelops the interior portions of the cerebral hemispheres which are composed of white matter and masses of gray matter constituting various nuclei and ganglia of the brain, including the thalamus.

The white matter beneath the cortex consists of three main fiber systems: *projection fibers* (ascending and descending), *association fibers* and *commissural fibers.* The majority of the *ascending projection fibers* carry impulses from sensory relay stations in the thalamus to the various sensory areas of the cortex, but some of them interconnect other sub-cortical ganglia and the cortex. Some of the *descending projection fibers* connect with sub-cortical centers of the brain and the cerebellum and impose a cortical contolling influence on them, but many of the descending fibers, originating in the motor regions of the cortex, carry motor impulses directly to centers in the spinal cord and are known as *pyramidal tracts.* The *association fibers* interconnect the different areas of the cortex and co-ordinate their activities. The *commissural fibers* (corpus callosum and anterior and posterior commissures) connect the two cerebral hemispheres.

The *cerebral cortex* in man is composed of six cellular layers which in most regions may be fairly well differentiated histologically by the size, shape and density of the cellular structures. Different areas of the cortex also show a varying histological pattern which has been used to demarcate various

structural subdivisions. Functional differentiations have not as yet been precisely determined for each of the structural subdivisions. Grossly, on the basis of both structural and functional considerations, the surface of the cerebral cortex may be divided into four major regions which are separated by deep furrows known as fissures. These regions are called the *occipital, parietal, temporal,* and *frontal lobes.*

The primary reception areas for *vision* are in the *occipital lobes* of each hemisphere; those for *audition* are in the *temporal lobes;* and those for general *somatic sensibility* (touch, pressure, pain, temperature and deep sensation) are in the *parietal lobes.* The arrival platforms for sensory impulses in each of these regions have been fairly well delimited and are small compared to the total area composing each of the lobes. The region immediately adjacent to the primary reception areas in each of these lobes is concerned with the elaboration of sensory messages and serves therefore for finer perceptual discriminations. Much of the area of each lobe is concerned with relating and integrating its activities with those of other lobes and is therefore known as an association area. This is particularly true of the region joining the occipital, parietal and temporal lobes and of the anterior part of the frontal lobe.

The *frontal lobe,* largest of the four lobes in man, lies anterior to a deep groove, the central fissure or fissure of Rolando, which extends laterally and anteriorly from a point at the top of each hemisphere in such a manner as to practically separate the anterior from the posterior half of the cerebral cortex. The posterior half of the frontal lobe is associated primarily with motor functions; the anterior half is a general association field which is responsible for some of the highest forms of integration and control in the cortex. Area 4 (see diagram in Fig. 20), part of which is concealed in the Rolandic fissure, is the primary motor area and gives rise to fibers of the pyramidal tract. Electrical stimulation of points along this area gives rise to movements of various muscle groups. For example, stimulation of area 4 at the top of the Rolandic fissure near the mid-line separating the two hemispheres causes contraction of muscles

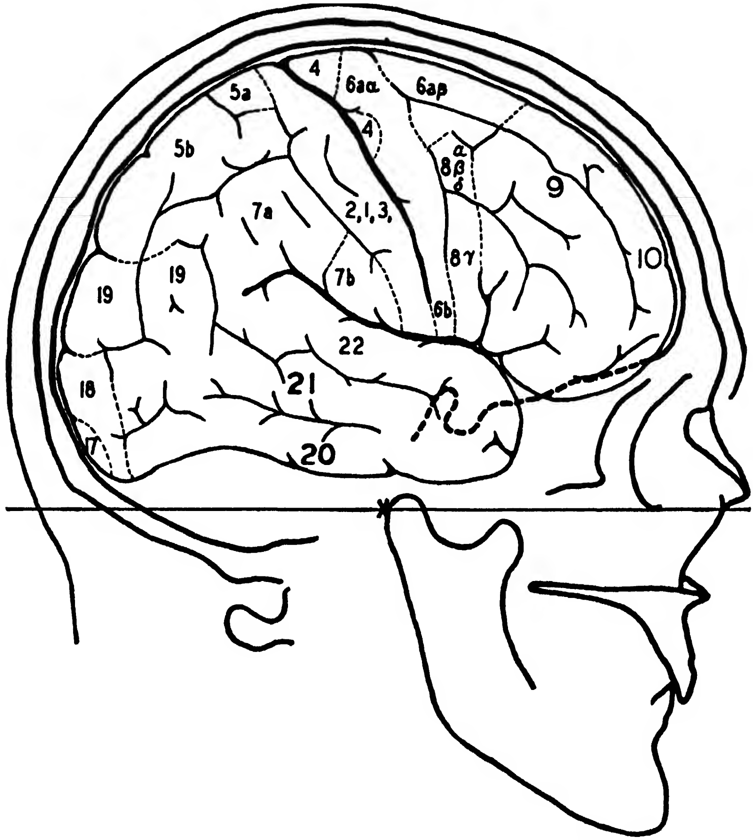


FIG. 20. A tracing of an X-ray photograph of the head of an adult male. A diagrammatic map of the brain adapted from Foerster and Penfield has been fitted to the skull. The four major lobes of the brain include the following numbered fields and others. Occipital lobe: 17 (primary visual field), 18 and 19; Parietal lobe: 3, 1 and 2 (common somatic sensory area), 5 and 7; Temporal lobe: 22 (auditory), 21 and 20; Frontal lobe: 4 (primary motor), 6 and 8 (pre-motor), 9 and 10 (frontal association areas). The central sulcus or fissure of Rolando is emphasized by a heavy black line separating areas 4 and 6 (motor) from 3, 1 and 2 (sensory); the lateral sulcus or fissure of Sylvius is emphasized by heavy black line separating areas 6, 7 and 8 from 22.

of the lower limb on the opposite side of the body; farther down stimulation causes reactions of trunk musculature; and still farther down along the fissure on the lateral surface of the cortex stimulation causes activity in the muscles of the head, face and neck. Area 6, known as the premotor area, is concerned with finer and more complex motor adjustments. It exercises some control over centers in the basal ganglia which are involved in postural adjustments and over others in the hypothalamus which are associated with activities of the autonomic nervous system.

In higher vertebrates, such as the cat and dog, the subcortical centers are capable of maintaining the vegetative functions which enable the animal to stand, walk, eat and so forth. It has also been demonstrated that certain forms of sensory discrimination and elementary forms of conditioning or learning may take place in the absence of the cerebral cortex. For the most part, however, the more complex mental functions require the cerebral cortex, and some of the higher functions in man are limited exclusively to the highly organized cortex.

METHODS OF INVESTIGATING CORTICAL FUNCTION

A number of types of investigation have been employed to determine the gross as well as the specific functions of the cortex. Most of the studies, other than the anatomico-histological, may be grouped under one of four headings with respect to method of approach: *Ablation*, *Stimulation*, *Clinical-Pathological*, and *Electroencephalographic*. Lack of space prevents description of experiments which illustrate each of these fields of investigation of cortical function. Therefore only a brief sketch of the method of approach and the nature of the results will be given for the first three; the electroencephalographic method and some of the results obtained by its use will be described in more detail in the next chapter.

Ablation experiments. There is a long history of experiments of this kind in which removal or inactivation of cortical tissue (either by excision, cautery, or more recently by thermocoagulation) in various areas has been correlated in animals

and human beings with the loss of certain sensory or motor functions. The removal of a small area of cortical tissue in the arm area of the motor region (area 4) in man or in an adult chimpanzee will produce a permanent paralysis of the arm on the other side of the body; removal of a corresponding area or even the whole motor field in one hemisphere of a lower vertebrate such as the rat will not produce a permanent paralysis of the opposite fore-leg. Similar differences between species within the vertebrate series have been demonstrated for sensory functions. In the lower vertebrates (fishes and amphibia) the visual centers are sub-cortical. Lashley has shown in the rat that ablation of the visual cortical areas abolishes pattern or form discrimination but that brightness discrimination is retained. Marquis has demonstrated in the dog, and Marquis and Hilgard have shown in the monkey, that destruction of the visual cortical areas does not abolish brightness discrimination. Yet in man, clinical observations have shown that bilateral lesions of the area striata (primary visual field) abolish even light discriminations. This difference in the corticalization of functions which exists among the species is known as the *principle of encephalization*, which means that with the progressive development of the cortex, phylogenetically, functions once mediated by lower centers are shifted to higher centers.

During the last twenty-five years, beginning principally with Franz's work on the monkey and Lashley's work on the rat, the destruction of cortical tissue has been related more directly to psychological functions such as perception, memory, learning and "intelligence." With the development of neurosurgery in recent years, the rather extensive ablations for the removal of brain tumors and other pathological tissues have provided opportunity for the study in man of changes produced in the higher psychological functions.

In addition to the brightness and form discrimination functions in the rat, Lashley has studied the effect of cortical lesions on the ability to learn and retain habits involved in solving a maze problem. Two particularly important concepts of corti-

cal function in the rat evolved from this work. One, the *law of mass action,* holds that, regardless of the areas involved, the amount of deficit or loss in the maze learning function is roughly proportional to the amount of cortical tissue destroyed. The fact that one cortical area is as important as another for the establishment and retention of a pattern of response such as is required for maze learning in the rat led to the second concept, *equipotentiality of function.* These findings led Lashley to conclude that integration in the cortex, at least in the rat, is dependent upon some sort of dynamic pattern of relationships rather than upon specific structural units.

The effect of ablations of various cortical areas in monkeys and chimpanzees has been studied, particularly by Jacobsen and by Klüver, in relation to problem solving. In general the results have indicated that in these primates the principles of mass action and equipotentiality do not hold, except within the limits of specific cortical areas involved in the function. This again is in agreement with the principle of encephalization of function, and is further indication that results obtained in one species are not always transferable to another. Most certainly results of cortical function studies in sub-human species may be generalized to include man only with great caution, for in man there is a higher degree of corticalization of function and consequently a greater specificity of function.

Jacobsen's studies of frontal lobe function in monkeys and chimpanzees are particularly important. He has found that ablation of the frontal association area in one hemisphere only had no effect on their performance of adaptive behavior tests. Bilateral lesions of the frontal association areas on the other hand did have a marked effect on their performance of these tests. In addition to restlessness, stereotyped behavior and changes in emotional responsiveness, he found that the principal defect in the performances on the tests resided in the animals' inability to integrate successive experiences in the situation over a brief period of time. In other words, it appeared that there was a defect in recent memory. This was brought out particularly in *delayed response tests* in which the

animal saw the experimenter place a piece of food under one of two or more cups. The cups were then shielded from the animal's view by a screen for varying intervals of time, after which it was removed and the animal allowed to respond. Whereas prior to the operation correct choices were made after intervals of delay ranging up to five minutes, animals with extensive bilateral lesions of frontal association areas were often unable to make correct responses after a five-second delay interval. Other tests in which the same type of defect could be demonstrated were: a *stick joining test,* in which two or more sticks had to be observed in different places about the cage and joined together in order to reach food on a platform; and a *problem box test,* in which food inside the box could be obtained only by pressing or pulling levers in a particular sequence.

Some other important studies of the effect of ablations in different cortical areas on various behavior functions are those by Klüver, who used some unique tests in the study of occipital, temporal and frontal lobe lesions in monkeys; and those by K. U. Smith, who studied various types of visual reactions in cats with occipital lobe lesions. Wiley and also Pennington have studied the responses of rats to auditory stimulation after destructive lesions were produced in the auditory area. "Reasoning" and other types of "intelligent" behavior were studied by Maier after various types of lesions in rats.

Dusser de Barenne and his collaborators developed a method of *thermocoagulation,* by which it is possible to control the depth of destruction of cortical tissue in a localized region. This has enabled them to investigate the function of various layers of the cortex. Pennington has more recently used the same technique.

Some of the most carefully studied human cases in which extensive ablations were made by neurosurgeons for the removal of tumors and so forth have been reported by Foerster, Penfield and Evans, Brickner, Dandy, and German and Fox. Brickner made rather extensive observations over a period of time of the behavior of a man whose frontal lobes were re-

moved by Dandy. Although for a while afterward he was able to carry on his work as a broker with a seat on the stock exchange, defects in recent memory, emotional instability, and lack of ability to synthesize or follow logical lines of thought in abstract terms proved to be serious handicaps. There were also rather distinct changes in the man's personality that were noted especially by his family. Among other things he became boastful, talkative, restless, and uninhibited, particularly in social situations calling for some restraint. Recent memory was particularly affected, but remote or old memories remained.

Some of the other reported cases of bilateral frontal lobectomy have also indicated changes in the affective or emotional life, defects in recent memory, and difficulties in synthesis of ideas. However, Penfield has reported a case of bilateral removal in which few, if any, of these symptoms were present after the operation.

Stimulation experiments. One of the earliest forms of investigation of cortical function began with the use of electrical stimulation of localized areas by Fritsch and Hitzig in 1870. They found that an electrical stimulus applied to points in the region of the cortex now known as the motor area in dogs and cats produced movements of parts of the body on the opposite side. The work of subsequent experimenters, particularly in Sherrington's laboratory at Oxford, provided detailed maps of the excitable points and the responses which were elicited by their stimulation; these workers also determined the limitations of the responsive motor areas.

Neurosurgeons such as Foerster, Penfield, Bucy and others have more recently mapped in detail the stimulable areas of the cortex in man and have noted particularly the effect of stimulation of the primary sensory and motor fields. Penfield and his associates have performed a number of brain operations under local anesthesia thus leaving the patient in a conscious state and able to report directly his psychological experiences resulting from the stimulation of various points on the cortex. In this way they were able to map the points which gave a sensory experience in the patient. Most of these points

were confined to the area of the postcentral convolution (cortical areas 1, 2 and 3 in diagram of Fig. 20), just behind the Rolandic fissure; however some sensory experiences could be elicited by stimulation immediately anterior to the fissure in the motor regions. The sensory experiences described by the patients consisted of numbness, tingling, and occasionally warmth and cold. Earlier, Cushing, the famous brain surgeon, performed stimulation experiments on patients who were conscious and found that points stimulated near the top of the postcentral convolution produced sensations localized in the lower limbs; stimulation of points lower on the convolution produced sensory experiences localized in higher parts of the body. Thus from stimulation of the sensory and motor areas on opposite sides of the Rolandic fissure it was found that there was a close correspondence between the level stimulated on either side and the region of the body in which the sensory experience was localized or the motor response observed. In both instances the effects were localized on the opposite side of the body; thus the right hemisphere of the brain is associated with sensory and motor functions on the left side of the body and vice versa.

Experiments by Dusser de Barenne and associates and by Bremer have utilized a different type of stimulation procedure, namely, the local application of an excitant drug, *strychnine*, which in a certain concentration "fires" or discharges the nerve cells in that area. When applied to a local region of the cortex in diluted form on a few square millimeters of blotting paper the strychnine heightens the excitability of the cortical cells (i.e., lowers their threshold of stimulation) to such an extent as to produce hypersensitivity to touch, pain and other sensory stimuli applied to the parts of the body that correspond to the strychninized area of the cortex. By observing the reactions of animals, when a hypersensitive area is stimulated, it is possible by this method to map the sensory areas of the cortex. More recently, Dusser de Barenne and McCulloch have used this method in conjunction with the recording of electrical responses from the cortex in order to determine the relation-

ships and interconnections between adjacent areas in the sensory and motor cortical regions. The results indicate that impulses are transmitted from sensory to motor areas and vice versa, but according to particular and complex patterns.

Clinical and pathological studies. The effects of pathological processes in the brain, especially those affecting the cortex (tumors and other abnormal growths; infectious processes such as encephalitis, meningitis and syphilis; toxic or poisoned states of the tissues; lesions produced by circulatory failures; injuries, atrophies and so forth) have been observed and reported by clinical neurologists and neuropathologists for many years. Often however there are many difficulties involved which prevent adequate study of such cases. The clinician is often too busy with a large number of patients to have time to follow in detail the changes which occur from day to day, or those that occur more slowly over a period of years in patients with brain pathology. Patients who have been studied carefully so far as symptoms and behavior are concerned may survive for years before obliging with the autopsy material necessary for the correlation of the effects with the precise anatomical lesion. Often the lesions are not limited to specific structures, as they may be in animal experiments, and thus interpretation becomes complicated. In some instances relatives refuse permission to perform an autopsy and to section the brain for study. As a result of a combination of these and other difficulties descriptions of behavior changes associated with specific types of brain lesions accumulate slowly in the literature. Summarizing these reports is often difficult because different methods have been used.

Perhaps a more serious criticism of clinical-pathological studies of cortical function may be directed toward the lack of adequate techniques and measuring devices for evaluating both behavior changes and loss or deterioration of psychological or higher mental functions, such as intelligence, personality, learning ability, memory, the perception of relations, and various psychomotor functions. Head, Goldstein and a few other neurologists and psychiatrists have sought to evaluate

losses in some of these higher mental functions by ingenious tests or techniques. Head's work on aphasia and other cortical disturbances was monumental. Goldstein has shown that frontal lobe lesions are accompanied by a loss of abstract ability, while at the same time there is retention of certain concrete aspects of behavior and thinking.' Goldstein has also stressed, in connection with lesions in any part of the brain, not only that the functions associated with the particular area destroyed are affected but also that psychological abnormalities in the total functioning of the cortex may result due to the influence of the specific part upon the whole.

In recent years a few psychologists trained in clinical and experimental methods have found a significant place in this type of work, and undoubtedly there will be more opportunities for the psychologist to collaborate with the neurologist, neurosurgeon or psychiatrist in the future. Babcock and Shipley have developed techniques for measuring deterioration of intellectual functions which are useful for determining the amount of certain functions lost due to progressive lesions or to operative removal of cortical tissue. Beck, Piotrowski and M. R. Harrower Erickson, working in conjunction with neurosurgeons, have made use of the Rorschach method to evaluate changes in personality introduced by cortical lesions and ablations. Halstead, working with neurological and neurosurgical patients, has developed tests or techniques based on grouping behavior, that is, the classifying and grouping of a large number of objects, which show up defects in the ability to perceive visual and other relationships. The particular type of defect, he attempts to relate to the locus and nature of the lesion. Hebb and others have reported on some of the aspects of intelligence which are affected by the removal of cortical tissue in various areas.

Electroencephalographic studies. The recording of electrical potentials from the brain in human subjects has developed within the last twelve years. In principle the method is similar to that described in an earlier section for recording the electrical variations in sensory and motor nerves and muscles. The

minute variations of electrical potential, measuring only a few millionths of a volt, are recorded from the surface of the scalp over various regions of the brain. Details of the method and some of the results from animal and human experiments will be described in the next chapter.

An advantage of this method over some of those just described is that, in normal human subjects, it provides an objective record of the electrical activities in an intact brain free from lesions, pathological defects and the depressing effects of anesthetics. There are, however, still many difficulties in the interpretation of electroencephalographic results and in the relating of them to behavior and psychological processes represented in cortical activities. Already, however, the method has provided some new concepts of cortical function. Not the least of these is the fact that in the waking state the cortex of the brain is "spontaneously" and continuously active electrically. This is due in part, apparently, to autonomous activity of the cortical cells. Moreover, the electrical activity is constantly in a state of flux or change, although it does maintain certain organized or patterned characteristics.

SUMMARY

In this chapter four methods of studying cortical function and some of the results obtained have been described. Of these the *ablation* and *stimulation methods* have contributed the bulk of our knowledge of cortical function both in animals and human subjects. By these methods sensory and motor fields of the cortex have been delimited and the nature of many of the functions subserved by each determined. In addition they have thrown light on some of the higher integrative functions of the cortex, such as those of the frontal association areas. Ablation and stimulation methods have also played a prominent part in helping to establish the principle of encephalization or corticalization of function which has to do with the progressive shift of control of functions from lower to higher neural levels in the phylogenetic development of the nervous system. The *clinical-pathological method* has made pos-

sible the correlation of many kinds of behavior deficit with cortical lesions in man, and has contributed many important cues for experimental work by means of other methods. The *electroencephalographic method* is new and has not yet demonstrated its full worth, but study of the electrical activity of the cortex appears to have considerable potential value in relating functional states of the intact brain to behavior processes.

CHAPTER II

Electroencephalography ("Brain Waves")

ELECTROENCEPHALOGRAPHY or the method of recording electrical potentials from the brain is relatively new, especially in human subjects. Caton, in 1875, discovered that it was possible to record electrical changes from the exposed brain of animals and several other investigators later confirmed this. The phenomenon attracted little attention, however, until 1929 when Berger, a German neuropsychiatrist, published the first report of studies of the electrical activity of the human brain. Since that time many workers both in this country and abroad have conducted investigations of the phenomenon in animals and human subjects. At the present time the number of published studies extends into the hundreds and the titles bearing on the electroencephalogram in human subjects alone total more than one hundred.

Following his initial report Berger published a series of studies in which he described in considerable detail many of the characteristics of the human electroencephalogram in normal and pathological conditions. Although his results were first looked upon rather skeptically, most of them were soon confirmed and in many instances extended by European and American investigators. The most extensive reviews of the subject at present are those by Jasper,[1] Davis,[2] and Walter.[3]

[1] Jasper, H. H., "Electrical Signs of Cortical Activity," *Psychol. Bull.*, 1937, *34*, 411-481.

[2] Davis, H., "The Electroencephalogram," *Tabulae Biologicae*, 1938, *16*, 116-131.

[3] Walter, W. G., "The Technique and Application of Electroencephalography," *J. Neurol. and Psychiat.*, 1938, *1*, 359-385.

THE NATURE OF THE ELECTRICAL ACTIVITY OF THE BRAIN

Despite the fact that many studies of the electrical activity of the brain have been made, both in animals and human subjects, the fundamental nature or basis of the phenomenon is still largely unknown. Most of the studies of the electroencephalogram in human subjects have been directed toward finding relationships between certain measurable aspects of the electrical activity and various physiological and psychological variables. Studies in animals on the other hand have attempted to determine the source of the potentials and to seek out the basic mechanisms involved.

As was mentioned in the preceding chapter, one of the most striking aspects of the electrical activity of the brain is its seemingly continuous, rhythmic flow and its apparently spontaneous, self-initiated character. Not only are these oscillating potential variations found in the cortex, but they are also present in sub-cortical masses of gray matter where there are aggregations of nerve cell bodies. In fact Adrian and his associates in 1931 demonstrated that slow, rhythmic potential waves could be recorded from groups of nerve cells in the isolated ganglia of the water beetle and in the isolated brain stem of the goldfish. In these instances it was shown that the potentials were independent of sensory impulses or of impulses from other regions of the nervous system. This seemed to indicate a *spontaneous* and more or less *autonomous* origin of the potentials in groups of associated nerve cells, apparently due to metabolic processes which are determined by the chemical environment of the cells.

In general it is believed that the potentials arise in the nerve cell bodies and are associated with rhythmic fluctuations of the excitability or irritability of the nerve cells. It appears that in some groups of cells the rhythmic changes of potential may be autonomous or self-governed as Adrian's experiments suggest, but in other groups only partly so; for studies by Bishop and his collaborators indicate that the rhythmic fluctuations of potential in cortical regions are partially, if not wholly, controlled

by the thalamus through "loop circuits" between the thalamus and the cortex. It is possible that a sub-group within a larger group of cortical cells, by virtue of a lower threshold of irritability, serves as a rhythmic "pace-maker" for synchronizing the activity of the whole group. It is apparent that some type of synchronizing mechanism or "pace-maker," either in the cortex or in sub-cortical regions (notably, in the thalamus), is necessary since the magnitude and pattern of the potential waves suggest that summation of potentials from many cellular units occurs rhythmically. This point of view has been proposed by Adrian particularly to account for the alpha rhythm (see below) of the electroencephalogram.

TERMINOLOGY

Berger originally recorded the electrical potentials from the surface of the scalp of human subjects by leading off the potential variations from electrodes attached to the front and the back of the head. The record of this electrical activity he called an *electroencephalogram* (abbreviated EEG), which means a tracing or graph of the electrical variations from the encephalon or brain. Sometimes the electrical changes are referred to as *brain potentials,* and in popular parlance the rhythmic wave-like tracings are frequently called *brain waves.* When records are obtained from local regions of the brain they are identified as occipital EEGs or frontal EEGs, depending upon the region from over which they are recorded. Records from the exposed surfaces of the brain in animals or in human subjects during operations are called *electrograms* to differentiate them from records obtained from outside the scalp and skull. If the electrograms are recorded from different regions of the brain they are identified as *cortical electrograms,* or *thalamic electrograms* and so forth.

Types of waves in the EEG. Berger identified two types of waves in the electroencephalogram of human subjects. The larger more prominent waves, of which he found evidence all over the skull when the subject was at rest, with eyes closed or in a dark room, he called *alpha waves.* These in normal adults

consist of rhythmic oscillations of about ten per second and are often referred to as the *alpha rhythm.* The other waves described by Berger were the *beta waves.* These were described as much smaller and faster waves, often superimposed on the larger alpha waves. Other types of waves or rhythms have been described, notably waves of lower frequency than the alpha rhythm and usually of greater magnitude. These have been called *delta waves* and are primarily associated with abnormal conditions in the waking state, although similar waves occur in normal persons during sleep or unconsciousness. In view of the fact that the electrical activity of the brain as recorded from the surface of the scalp may consist of rhythmic variations ranging from less than one per second to fifty or more per second, it appears wise to identify other types of waves by their frequency and amplitude limits, or in some cases by their form or pattern.

METHODOLOGY

Since the method and procedure are in general similar for all studies, at least with humans, it will be described here before a survey of the results of such studies is given. Figure 21 illustrates the apparatus and method.

Electrodes and their placement. For recording EEGs from the surface of the scalp of intact human subjects, the particular type of electrode used is not especially important so long as it makes satisfactory contact with the scalp through an electrode jelly or some other electrolyte solution. Many laboratories use small, flattened or cup-shaped pellets of solder about 5 to 10 millimeters in diameter. Others use silver electrodes which may be made relatively non-polarizable. A fine, flexible, insulated wire is attached to each electrode and leads to the amplifiers or a switch box where various combinations of interconnections between electrodes may be made.

No standard system of electrode placement is followed by all workers at present, although in general there is similarity of the electrode arrangement on the surface of the head in order that samples of the electrical activity may be obtained from the

various representative head regions. The exact position of the electrodes on the head is measured with respect to certain well-defined anatomical landmarks such as the *glabella,* the region between and at the level of the eyebrows, and the *inion,* a small

FIG. 21. Apparatus and procedure for recording brain potentials. Switch-box (*A*) is set to record electrical potentials from the various pairs of electrodes attached to the surface of the scalp and the ear lobes of the subject (*B*). The wires from the head are attached to the appropriate amplifier connections at the plug-in-box (*C*). The potentials are conducted to the voltage amplifiers (*D*) and the power amplifiers (*E*). The amplified potentials (magnified several hundred thousand times) activate the four-pen Grass inkwriting oscillograph (*F*) or the Westinghouse four-element mirror oscillograph (*G*). The inkwriting oscillograph traces the electrical variations from the brain as waves on the folding paper tape; the mirror oscillograph reflects a beam of light which traces them on moving photographic paper in camera (*H*). (Lindsley, 1941.)

bump or occipital protuberance at the back of the head; on the side the measurements may be made with respect to the auditory meatus, or ear canal.

Two systems of recording are often described by workers in different laboratories; one, a "monopolar" or "unipolar" system, in which one electrode ("active") is on the surface of the scalp and the other grounded electrode ("indifferent" or "in-

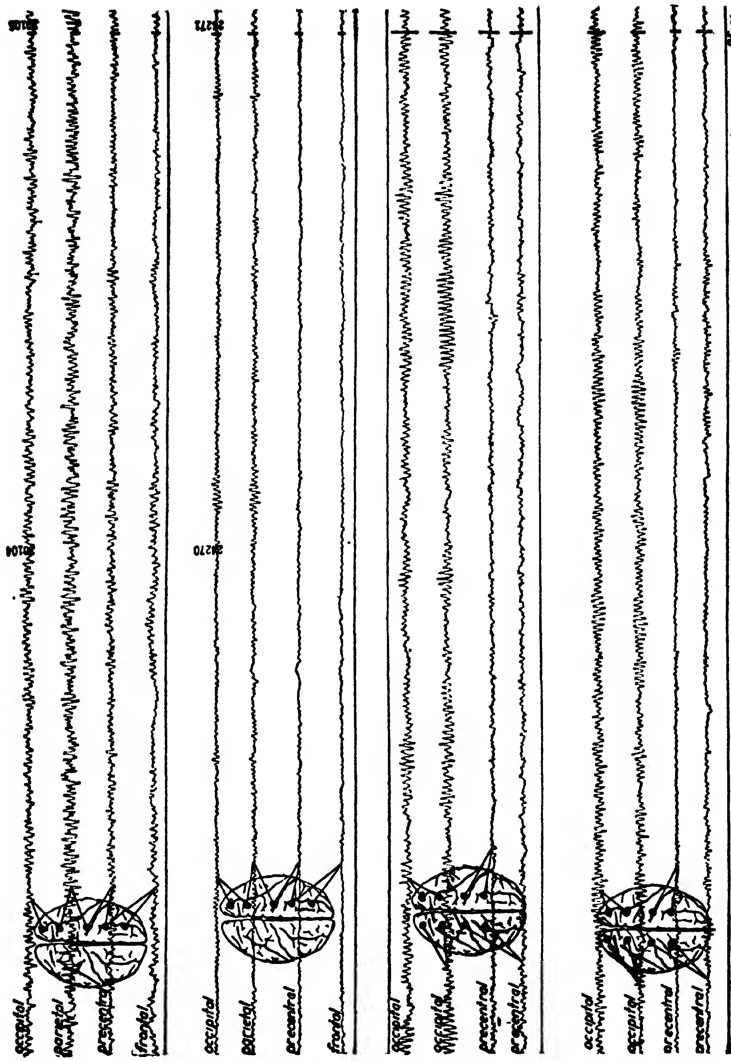


FIG. 22. Normal EEG records from four different adult subjects, showing individual differences in pattern. The top record is from a 32-year-old, well-adjusted married woman. The second record is from a young physician, 27 years of age. The third record is from a young college man, aged 21. The bottom record is that of a young man, aged 20, who was convicted of murder. In all of these records note the difference in pattern of activity in different regions of the head. Alpha waves are the prominent rhythmic waves appearing strongest in the occipital and parietal regions. Beta waves are fine, fast waves best seen in the frontal and precentral regions. (Lindsley, 1940.)

active") over a relatively inactive region, such as the lobe of the ear or the mastoid process directly behind the ear; the other system, known as "bipolar," consists of two electrodes (both "active") spaced perhaps two or more inches apart over active regions of the brain, neither electrode being grounded. There are advantages and disadvantages of both systems and most workers make use of both for certain purposes. A bipolar arrangement of electrodes is illustrated in Fig. 22.

Amplifiers and recording apparatus. High-quality amplifiers of special design are required. They must have sufficient amplification to magnify the minute brain potentials of a few millionths of a volt as much as 500,000-fold or more. The amplifiers must be capable of amplifying, without distortion, oscillations of potential ranging in frequency from less than one per second to more than 60 per second. The over-all response of the amplifying and recording system should be flat, with a constant input voltage, over the range of frequencies measured.

Various types of recording units have been used. Berger's first studies were made with a string galvanometer. Most laboratories use some form of oscillograph, either a mirror or galvanometer type oscillograph, a cathode ray oscillograph, or an inkwriting (crystal or electro-magnetic) oscillograph. Because of its economy and convenience the inkwriting oscillograph is employed in most laboratories for studies of the human EEG. Long, slender, tubular pens trace the potential changes in ink on paper tape moving at a uniform speed. From one to six or more of these inkwriting units may be arranged to trace records from as many regions of the head on a wide paper tape. An example of the simultaneous tracings from four regions of the head by means of a four-channel amplifying and recording system is shown in Figs. 22 and 23.

Procedure. After the desired locus for the electrodes has been determined by measurements a few hairs are cut at each spot to allow the electrode to fit snugly against the scalp. The scalp is cleaned with acetone or an alcohol-ether solution to remove sebaceous secretions or oils. The electrode, with a small

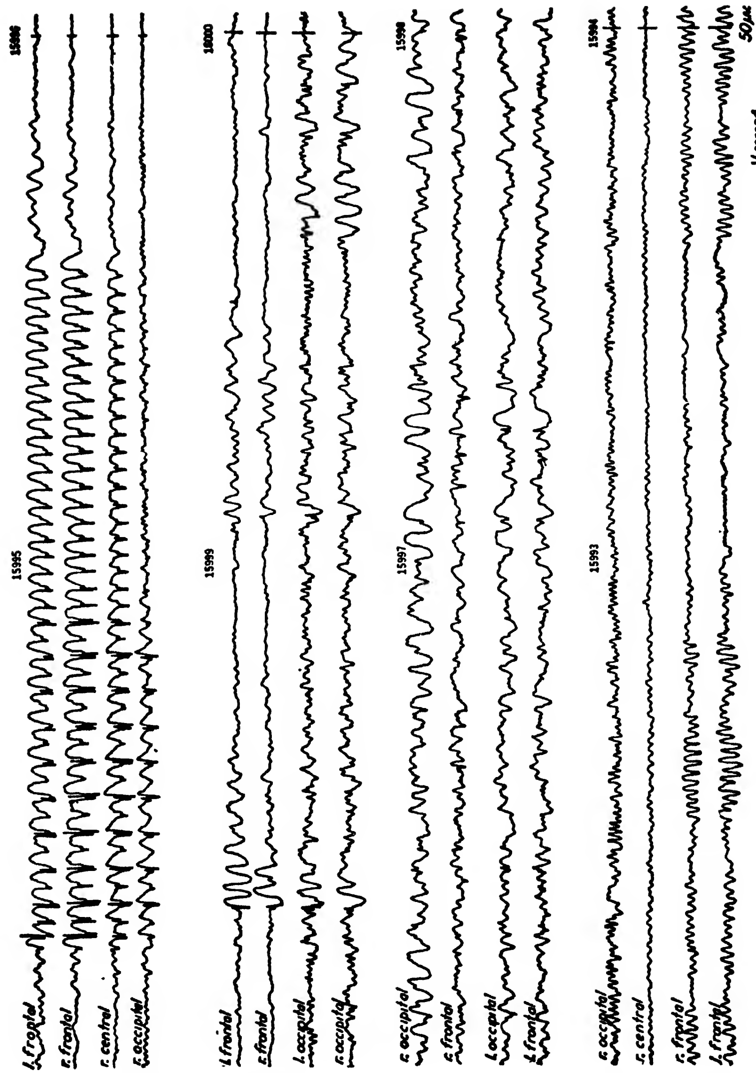


FIG. 23. Abnormal EEGs. The top record from a 7-year-old boy with epilepsy shows a typical spike and slow wave pattern. The petit mal seizure responsible for this series of spike and slow waves was induced by about 30 seconds of overbreathing. The second record is that of a 10-year-old boy recovering from a severe attack of encephalitis. The third record is from a 21-months-old girl with a diagnosis of porencephaly (failure of cortex to develop in a large portion of left hemisphere). The bottom record is from a 10-year-old boy whose behavior difficulties consist of hyperactivity, restlessness, unmanageability, facial tics, and the making of queer noises. Compare these records with the normal ones in Figure 22. (Lindsley, 1940.)

amount of the electrolyte jelly on its under surface, is then cemented in place with collodion. When all electrodes are in place the subject is asked to lie quietly and relaxed on a cot in an electrically shielded and relatively soundproof room so as to be free from external stimulation. Either the room is made dark or the subject closes his eyes, for visual stimulation disturbs the pattern of activity from the brain.

Records are usually obtained from various combinations of regions during a period of fifteen to twenty minutes or more. This provides adequate samples of record from all regions of the head and at different times during the recording period. Ordinarily to make a careful electroencephalographic examination of a subject at least one hour is required, including the time spent in putting on and taking off 10 to 16 electrodes and in obtaining a systematic series of records. Making a detailed analysis of the records requires considerably more time.

The chief quantitative aspects of the analysis consist of measuring the average number of waves per second (*frequency*), the average height or *amplitude* (converted into microvolts as a measure of magnitude), and the amount of time that waves of a certain frequency are present (*per cent time present* or *alpha index* if referring to the amount of alpha waves present). Other factors such as the form of pattern of the waves are also considered, but are not readily measured.

CHARACTERISTICS OF THE EEG IN NORMAL HUMAN SUBJECTS

Resting state. In normal adults, relaxed and at rest in the absence of stimulation, two types of waves appear in the record of electrical activity from the brain. The most prominent aspect of the records from any region of the head are the *alpha waves*, or *alpha rhythm*. These waves in adults have an average frequency of about 10 per second. In magnitude they range from just barely measurable waves of one or two microvolts (millionths of a volt) to perhaps 100 microvolts, although usually they average considerably less than 50 microvolts. The *beta waves* are much smaller and faster waves. They were originally described by Berger as ranging in frequency from about 20 to

50 per second. The most prominent waves of this type are usually found over the motor region of the brain and range in frequency from about 18 to 30 per second. These small waves are not very prominent in the posterior regions of the head but are sometimes superimposed on the alpha waves which are usually stronger in these regions than in the anterior regions. The small size of the beta waves and their higher frequency make them more difficult to study and sometimes they are confused with muscle action currents arising from tense scalp muscles. The beta waves have so far failed to show very significant correlations with stimulation, motor activity or other factors, and will therefore be mentioned relatively little in the subsequent discussion of results.

Effect of stimulation on the alpha rhythm. When a subject is in a dark room with eyes open, the occipital EEG tracing may show a more or less continuous series of rhythmic alpha waves. If a light is turned on or a visual stimulus presented the alpha rhythm is either completely blocked or is greatly diminished in amplitude approximately one-third of a second after the onset of the light. If the light remains on, the alpha rhythm tends to remain blocked, although it may come and go at reduced amplitudes. If, however, the light is turned out, the alpha waves reappear in full force, or even slightly increased in amplitude and frequency, as if recovering from inhibition with a rebound, about 1 to 1.5 seconds later. An auditory stimulus such as a tone or a voice, or even a tactual stimulus, will also block the occipital alpha rhythm, but seldom as effectively as a visual stimulus. Usually after repeated auditory stimulation adaptation occurs and the alpha rhythm is no longer affected by the stimulus to the same degree as originally. If a subject is physically or emotionally tense on entering the experimental room, it is often found that the alpha rhythm is not as well developed as later when a relaxed condition is attained.

Variability of the alpha rhythm. In normal adults the alpha rhythm, though averaging about 10 per second, may range from 8 to 12 or 13 per second in different individuals. Likewise, the amplitude of the alpha waves in different indi-

viduals varies considerably. The percentage of time that the alpha rhythm is present in a given length of record, the alpha index, may also vary in different subjects, ranging from practically zero to 100 per cent. The average alpha index is in the neighborhood of 60 per cent. Finally, the pattern of the EEG in different individuals is often strikingly different. In one subject the alpha rhythm may be extremely constant, of uniform amplitude and of generally even, regular pattern, whereas in another the alpha waves are irregular and vary in amplitude and pattern. These differences between individuals have been the subject of considerable interest and speculation, especially since it has been found that the pattern of activities in relatives is more similar than that in non-relatives, and in identical twins the patterns and characteristics of the activity are very similar.

Differences within the individual, on the other hand, are not great. A particular individual tends to reproduce his pattern and other characteristics of the EEG with exceptionally little variation from day to day and even from year to year. Thus, although the EEG of a particular individual remains quite constant from time to time, the pattern of activity of the EEG for different individuals shows distinct differences (see Fig. 22). So far, however, these differences have not been found to be significantly related to differences in behavior characteristics, personality or other factors in normal subjects.

GROWTH OF THE ALPHA RHYTHM IN CHILDREN

Berger first called attention to the fact that the frequency of the alpha rhythm is lower in children than in adults and that it is absent in newborn infants. Later studies by Lindsley and by Smith on large numbers of children of varying ages ranging from newborn to adulthood have established certain developmental characteristics of the EEG. As is shown in Fig. 24, the *occipital alpha rhythm* is absent in a young infant during the first 3 to 4 months. The frequency of the waves at its onset in the third or fourth month is usually about 3 to 4 per second. By twelve months the frequency has increased to 5 or 6 per

second and by three years the average frequency is about 7 per second. The frequency continues to increase with age until the adult frequency level of 10 per second is reached, on the

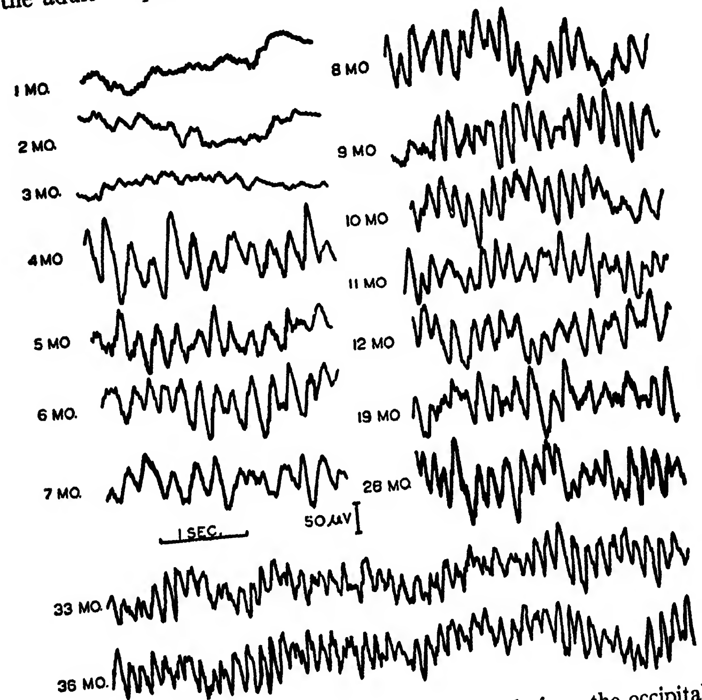


FIG. 24. Tracings of the brain potential records from the occipital region of a child during the first three years of life, showing the onset of the occipital alpha rhythm at four months (frequency 4.0 per second; amplitude 64 microvolts) and its subsequent development to three years of age (frequency 7.3 per second; amplitude 32 microvolts). (From Lindsley: *J. Genetic Psychol.*, 1939, 55, 197-213, by permission.)

average by children of about twelve years of age. The amplitude of the alpha waves grows during infancy but after about three years it declines steadily until adulthood, the decrease in amplitude being roughly proportional to the increase in frequency.

Smith's extensive studies of newborn infants and young children indicate that an alpha rhythm (7 per second) is present *over the sensory-motor regions* at birth. This rhythm is found mainly in newborn infants when they reach a quiescent, relaxed state, apparently just before going to sleep, but is not usually found during the first year in infants during the active, waking state. By the end of the first year, however, it is observed during the waking state and thereafter increases in frequency at about the same rate as the occipital alpha rhythm.

The onset of the *occipital alpha rhythm* at about 3 months of age appears to be associated with the functional and maturational development of the visual cortex and to some extent may be said to correlate with the development of visual habits of response to moving objects. The appearance of the *alpha rhythm* in the *sensory-motor regions* at birth is in agreement with the advanced maturational status of the cortex in those regions as compared with the occipital or temporal regions where the rhythm is absent at birth and the development of the cortex somewhat retarded. Eventually, the functional development of the cortex, as revealed by the EEG, may be found to be more precisely related to the cortical cellular development and to the functional capacities of the cortex as revealed by behavior indices.

PHYSIOLOGICAL VARIABLES AND THE EEG

A number of lines of evidence suggest that the alpha rhythm is associated with or affected by metabolic processes in the brain.

Temperature. Hoagland and his associates have demonstrated that raising the body temperature of subjects by diathermy treatment increases the frequency of the alpha waves.

Metabolism. The frequency of the alpha waves shows a significant relationship to metabolic rate and the administration of thyroid extract increases the frequency of the alpha rhythm.

Blood sugar. Reduction of the blood sugar level below a critical limit will produce coma and unconsciousness, but before

this stage is reached slowing of the alpha waves occurs and other slow-wave components appear in the EEG.

Acid-base balance. Changes in the pH of the blood (a measure of the hydrogen ion concentration, which determines its acidity or alkalinity) beyond certain limits have been reported by Gibbs and associates to alter the waves of the EEG. There is a slowing down of the waves for alkalinity, and a speeding up for acidity.

Hyperventilation or overbreathing. Deep, regular breathing for a number of minutes produces dizziness and eventually unconsciousness may ensue, but before this happens a normal individual may show evidence of abnormally slow waves. Ordinarily, normal subjects do not show abnormal manifestations in their EEGs within a period of three minutes of voluntary overbreathing, but in patients with epilepsy or even in individuals with epileptic tendencies only, the abnormal patterns of activity appear in the EEG often after less than one minute of overbreathing, and sometimes epileptic seizures occur. This method of overbreathing is frequently used to bring out latent signs of abnormality in the EEG in other types of patients.

Low oxygen. Breathing air mixtures which are only 8 to 10 per cent oxygen (normal air is 20 per cent oxygen) tends to slow the alpha rhythm and eventually, with the onset of dizziness and loss of consciousness, produces large, slow delta waves such as are found in sleep and in certain abnormal conditions.

PSYCHOLOGICAL VARIABLES AND THE EEG

The correlation between psychological variables and changes in the EEG has so far been largely incidental to other investigations, although a number of studies, particularly those by Jasper, Travis, Knott, and Williams, have been directed specifically toward certain psychological factors. Kreezer [4] has discussed a number of the implications in the EEG which have a bearing on psychological processes. Much of this type of in-

[4] Kreezer, G., "The Electro-encephalogram and Its Use in Psychology," *Amer. J. Psychol.*, 1938, *51*, 737-759.

vestigation has awaited a better understanding of the fundamental nature of the electrical phenomena of the brain.

Attention has already been called to the fact that *sensory stimulation,* visual, auditory and tactual, modify the EEG and block the alpha rhythm. Berger, Adrian, Jasper and others have pointed out that *attention* or *"set"* has something to do with the effectiveness of a stimulus in blocking the rhythm. Adrian also believed that form or pattern vision, that is, *perception,* was necessary for effective blocking of the rhythm by visual stimuli. Jasper and Cruickshank have shown that *visual after-images* are correlated with modifications of the EEG after a period of stimulation.

The depressant effect of *anticipation, emotional states,* and *anxieties* have been noted by a number of workers. Knott, Williams and Freeman have been concerned with the influence of *tensions, preparatory set, mental set* and so forth, both on the resting EEG and on the modifications of it by stimulation. Apparently the effects in a number of these instances are contingent upon conditions or levels of excitation in the nervous system and also upon factors which support or modify these conditions; namely, the influx of sensory impulses from the periphery. The fact that relaxation is necessary to the free flow of cortical alpha rhythms is one indication of this.

A variety of other psychological factors have been shown to have effects on, or to be related to, the EEG. Among these are "mental work," speech, silent and oral reading, flicker and fusion frequencies of visual stimuli, and hypnosis.

Intelligence and the EEG. So far no significant correlations have been found between the measurable aspects of the EEG and intelligence as measured by tests among groups of individuals of normal intellectual development or above. Kreezer, however, has reported some extensive investigations of the characteristics of the EEG in relation to the intelligence level among the feeble-minded. He has worked with adult feeble-minded of varying mental age levels and has thus in effect been able to hold chronological age constant (at least so far as its effect on the growth of the EEG is concerned). Among a group

of feeble-minded of the *Mongolian type* he has found just barely significant correlations between mental age and the alpha index and between mental age and the alpha amplitude. In *hereditary types* of feeble-mindedness he also found a relationship between mental age and alpha frequency which was barely significant. Other studies are under way, and in one of them personal communication to the author has indicated that significant relationships between frequency of the alpha waves and intelligence are being found. However, on the basis of published studies to date, one must conclude that variation in intelligence as we know it among the individuals of a normal population has no counterpart in the currently measurable aspects of the EEG.

THE ELECTROENCEPHALOGRAM DURING SLEEP

Loss of consciousness in sleep, under anesthesia, or as a consequence of other conditions, produces striking changes in the EEG. Waking patterns of activity are gradually abolished and new patterns substituted. The new patterns frequently consist of large, slow, delta waves which resemble waves found in certain pathological conditions such as brain tumors, epileptic conditions and so forth. The slow waves which occur in normal subjects each night during sleep would be considered as evidence of pathology if they occurred in the same individuals when awake. Davis, Davis, Loomis, Harvey and Hobart have made extensive investigations of the EEG in human subjects during various stages of sleep. The results are extremely interesting for they seem to indicate that the EEG is one of the best methods of measuring the level or depth of sleep. Not only are there different levels of sleep but there are various stages in going to sleep. In these studies, based on 28 experiments in 14 different subjects, the state of consciousness was followed by means of prearranged signals. If a subject realized that he had just "drifted or floated off" for a moment he squeezed a signal bulb once; if he felt that he had just awakened from "real sleep" he squeezed the bulb twice. These signals appeared on the continuous EEG record.

The five main stages of going to sleep and their EEG correlates are as follows: (1) Relaxed and resting, but awake—normal alpha rhythm present, but occasionally interrupted. (2) Distinctly drowsy but not yet asleep; tendency to float off in "light sleep" toward end of this stage—alpha rhythm drops out and low amplitude slow waves of 4 to 5 per second appear irregularly. (3) Real sleep—spindle-shaped volleys of 14 per second waves appear and the slow waves grow in magnitude and lessen in frequency; the slow waves occur randomly rather than in a definite rhythm. (4) Deeper sleep—longer and larger 14 per second spindles; the slow waves are larger, slower and more rhythmic. (5) Very deep sleep—14 per second spindles almost disappear; slow waves are very large and very slow and lack rhythm (random).

In going to sleep there is some tendency for the EEG patterns to fluctuate from one stage to another in the (1), (2), (3) range. Dreams occur chiefly during the (2) and (3) stages. As soon as the alpha waves disappear during the (2) stage, the subject's awareness of external stimulation is depressed. During deep sleep the slow waves may reach a magnitude of 500 microvolts (10 or more times the subject's normal waking alpha wave magnitude) and may have a frequency of one wave every 3 or 4 seconds.

THE EEG IN ABNORMAL AND PATHOLOGICAL CONDITIONS

Clinical application of the EEG has made rapid strides during the past few years. A great variety of neurologic and psychiatric conditions have been studied. Space does not permit a description of the many interesting results. Briefly, it may be said that the EEG has become a useful tool in aiding the neurologist and neurosurgeon in the diagnosis of certain types of neurological disorders. Its greatest usefulness lies in its value for the diagnosis and study of epileptic and related conditions. Certain characteristic patterns of wave formations appear in the epilepsies. For example, one well-known pattern has been called the spike and slow wave pattern. This wave formation tends to repeat itself at a frequency of about 3 per second and

the magnitude of the waves is much higher than normal. The spike and slow wave pattern is typical of petit mal epilepsy (see top record of Fig. 23). Other patterns are found in the grand mal and other types of epilepsy.

Another clinical application of the EEG is in the discrimination and localization of brain tumors and other pathological conditions in the cortex. The precise localization of a cortical tumor can frequently be made more accurately by the EEG than by other methods. Certain other psychiatric and neurologic conditions have been studied with some success. Figure 23 illustrates a few of the types of abnormal EEGs found. These may be contrasted with the normal records shown in Fig. 22. It is of interest to note that in children with behavior disorders, but without physical or neurological symptoms, there are frequently signs of abnormal function or cortical disturbance in their EEGs. The bottom record of Fig. 23 is an example of this.

SUMMARY

In this chapter a brief sketch of the methodology and some of the results of electroencephalographic studies have been presented, particularly results from studies of normal human subjects. No attempt has been made to describe the numerous studies of the electrical activity of the brain in animals, although a few of the salient findings were mentioned in connection with the discussion of the nature of the electrical activity of the brain. One of the most important aspects of the electrical activity of the brain as represented in the electroencephalogram is its spontaneous, autonomous, and continuously rhythmic character. Rhythmic alpha and beta waves make up the electroencephalogram of normal adult subjects. The rhythmic patterns of activity are subject to change by stimulation and a variety of physiological and psychological conditions of diagnostic value in clinical medicine. The EEGs differ from one individual to another in pattern but remain very much the same in one individual from time to time. The frequency of alpha waves changes with age up to about 12 years.

Part Four

COMPARATIVE PSYCHOLOGY

By

H. F. HARLOW

University of Wisconsin

CHAPTER 12

Animal Behavior

THERE can be no doubt that primitive man made many unrecorded observations on the behavior of lower animals, when, for reasons of dire necessity, he observed and studied the habits of the animals upon whom he preyed and who preyed upon him.

The first question of control of animal behavior was nip and tuck between the experimenter and the subject, with the experimenter taking the lead when he devised his snares for the hare and his pitfalls for the elephant.

Controlling the animal through knowledge of his habits attained greater significance with the advent of domestication, although here too there was at first a question as to whose habits were being molded. Even today the primitive Siberian tribes of the Chuckchee adjust their habits in large part to the movements of the reindeer.

Now, aside from the delight of satisfying sheer human curiosity, the behavior of the animal is studied and controlled for economic gain far above the subsistence level. But beyond this, there is a further reason for studying animal behavior and this is to gain information leading to the control of man's own behavior.

ANIMALS IN PSYCHOLOGICAL RESEARCH

Curiosity and desire for economic gain, immediate or remote, are in themselves legitimate motives for the study of animal behavior. The use of animals in psychological research, however, is motivated by the assumption that for certain problems, or for limited aspects of certain problems, *lower animals are better subjects than are human beings*. Their chief advantages to the experimental psychologist are discussed below.

1. The greater simplicity of the behavior of lower animals is frequently an aid in the analysis of complex functions.

This advantage is nicely illustrated in various problems of learning and forgetting. The vast literature on the conditioned response represents one attempt to solve the tremendously complex problem of the fundamental conditions involved in the formation of new behavior patterns resulting from experience. The attempt has been made by conditioned response techniques to determine the essential stimulus conditions for learning, to determine the precise effects of the frequency of stimulus presentations, to equate the influence of external distracting conditions, and to evaluate the relative influence of time and internal-external disturbances on the dissolution of newly formed reaction-patterns. From one point of view, human subjects are inferior to lower animals as experimental subjects in work of this type; for by the use of language they may repetitively stimulate themselves even after the physical stimulus has been removed. Because of the human capacity and tendency to talk, the experimenter never knows exactly what stimuli are producing the learning; the possibility of the use of implicit, subvocal speech prevents the experimenter from ever knowing how many times the stimuli have been implicitly presented. The capacity of human beings to refresh their memory by self-stimulation through language makes it extremely difficult to formulate exact laws governing learning and forgetting. Furthermore, learning by human subjects is greatly influenced by their "mental sets." Such sets as the intention of the subjects to try to learn, or not to try to learn, or to try not to learn, may entirely obscure the role played by the frequency of the intended paired stimulation or the role of distracting conditions.

Similar problems may confront the psychologist when working with lower animals, but all these problems are exaggerated in man as a result of the greater complexity of the processes involved in human learning, particularly those involving language.

2. Control of the external environmental conditions is possible to a greater degree with lower animals than with man.

An effective method for investigating the relative influence of heredity as opposed to environment, for instance, is to compare the effects of environments made as diverse as possible on individuals of identical, or at least similar, hereditary potentialities. To apply this method to a study of the procreative behavior of human beings, we might work with a number of pairs of identical twins. A group made up of one member from each pair might be raised from birth in a pitch black, soundproof room with constant humidity and temperature. Upon sexual maturity, the behavior of this group would be compared with that of a group composed of the co-twins, raised in a normal environment and given maximum opportunities to benefit from experience derived from years of traditional human culture.

Such an experiment, in all probability, will not and cannot be carried out with human subjects, but similar experiments have already been conducted with subhuman animals. Rats of the same litter have been reared in isolation from shortly after the opening of the eyes until sexual maturity. Under these circumstances the sexual act appears in complete form when pairs of subjects are first introduced into a single test chamber. With mature monkeys and chimpanzees, it is doubtful if these behavioral patterns would appear so perfectly if all previous experience had been denied. Unfortunately, these animals have never been reared in isolation to the point of sexual maturity. Such a technique is, however, entirely feasible and would provide a satisfactory answer for a problem which cannot be solved with human subjects.

Valuable knowledge concerning the nature of internal drives, the maturation of behavior patterns, and the interrelationships between heredity and environment have been disclosed by procedures which of necessity involved rigorous and often protracted control over the external environment of the subject.

3. The life span of most subhuman animals is shorter than the life span of man. Investigations of the inheritance of psychological processes and of the effect of long-time psychological and physiological influences on behavior may therefore be more

conveniently and efficiently studied in lower animals than in man.

If ample resources were available, would it be possible to rear a race of supermen with infinite wisdom, emotional stability, and physical perfection? If this is true, how many generations would it take?

Some answers to these problems might be available in ten to twenty generations, if we could build a wall around the state of Oklahoma and use all of its inhabitants for psychological study. Even if this were done, the experiment would take at least 1000 years and "we" would never know the answer. Rats, however, mature rapidly and several generations may be bred under rigidly controlled environmental circumstances in a single year. It is possible, therefore, for an experimental psychologist to attain a partial answer to such a problem by using some objective measure of intellectual ability, emotional stability, or motor co-ordination and by using the technique of selective breeding with his subjects. Furthermore, the scientist who started the experiment may finish it himself and thus gain more than posthumous fame. Even if such experiments can be eventually carried out on human subjects, data from lower animals will be extremely useful in pointing out techniques and difficulties intrinsic to any such long-time experiment with man as the experimental animal.

4. It is possible to make better controlled and more drastic alterations in the physiological state and organic structure of lower animals than it is in man.

Aristotle believed that our most complex behavior processes were regulated not from the brain but from the heart! Today we not only know that this conception is false, but we have available a truly vast amount of information concerning the intrinsic functions and interactions of many of the minute parts of the central nervous system. In large part, our present intellectual sophistication concerning brain functions has resulted from studies of animals lower in the phyletic series than man. It is possible to perform operations on rats, cats, dogs, monkeys, and chimpanzees in which precise, limited, and localized

portions of the brain are destroyed or inactivated. These subjects can be tested both before and after operation and the exact behavioral changes determined. Furthermore, subjects which are free from the debilitating effects of earlier illness can be selected.

It is true that our knowledge of brain functions is in part derived from human pathology, but clinical cases cannot provide a complete answer to such problems. Vascular accidents, tumors, gunshot wounds, and motorcycle accidents do produce destruction of brain tissue, but seldom, if ever, do these traumatic events *destroy the particular area and only the particular area* whose functions a scientist may desire to study.

LIMITATIONS OF PSYCHOLOGICAL RESEARCH ON SUB-HUMAN ANIMALS

No other animal is exactly like man, and data obtained from other animals are applicable to man only in so far as they relate to essential interspecies similarities. Thanks to comparative anatomy and comparative physiology we now have a great mass of data describing the ways in which man and lower animals resemble or differ from each other. These data serve as guides and keys, indicating uses and limitations of the data which deal with the comparative behavior of animals and man. Physiological research may be used by way of illustration.

The hormones of man and other animals are essentially similar. The remarkable advances made during the last two decades in the knowledge of hormonal function are dependent upon this fact; the functions of the endocrine glands may be studied experimentally in lower forms and the information there learned applied in large part to the human organism.

Transfer of information from animals to man is not always so successful. For example, dinitrophenol proved a safe and useful way of speeding up metabolism and reducing weight in the rat. The same results were obtained with man—but with a series of additional and unfortunate results including extreme nervousness, damage to the visual mechanisms, and even death.

In the study of psychology our fundamental interest con-

cerns processes primarily related to functions of the central nervous system; and the central nervous systems of man and lower animals differ more than almost any other bodily system. Although comparative anatomy and physiology have given us a wealth of information concerning these differences, the essential dissimilarities must never be ignored.

Certain forms of behavior are also predominantly or exclusively human; here the study of lower animals can supply either no data or only fragmentary data. Such phenomena include language (particularly speech), social organization, and the more recondite mental processes.

Nevertheless, animal research has been found useful in the entire psychological field, from the most isolated reflex to the study of complicated personality traits involving the organism as a whole.

In the following pages are described experiments bearing on three important problems: (1) the nature of complex mental processes in monkeys, (2) the inheritance of emotional patterns and of maze learning ability, and (3) social organization and cooperative behavior in primates.

COMPLEX MENTAL PROCESSES IN MONKEYS

In common parlance we speak of some problems as being easy and some as being difficult. We commonly assume that mental processes vary in their degree of complexity; complexity in terms of stimulus-response relationships or complexity in regard to the number and patterning of nervous pathways, without consideration of the exact meaning of these terms.

Complex and difficult problems frequently entail a larger number of isolated steps than problems whose solution is easy or simple. Yet this is by no means a complete answer, since the learning of a maze or the typing of a manuscript are tasks involving the integration of a large number of separate steps without demanding unusual ability. Stacking boxes one on top of another to obtain a food reward appears to demand considerable ability in spite of the fact that few behavioral steps are involved.

It has been suggested that the complexity of various mental processes might be differentiated in terms of the number of neurones involved or by the multiplicity of interacting neurone patterns. Such a differentiation is, however, pure conjecture beyond the known fact that gross cortical injuries affect various mental processes in a differential manner. Actually we have no knowledge of the number or pattern of neurons involved in any intellectual function since no technique has yet been developed that will allow us to follow the patterns of cortical activity during the learning process.

Probably the best criterion that we have for the separation of higher and lower mental abilities is the criterion of behavioral development broadly applied. This criterion assumes that the less complex behavior develops first and the more complex later, both in the development of the individual and in the development of the race. Thus as the child grows older he enjoys new behavioral potentialities of which he was previously incapable and this development appears in an orderly sequence. Likewise, as we ascend the phyletic scale a similar development appears.

The *flatworm* can learn a simple conditioned response, but the acquired response is never stable and is rapidly forgotten.

The *goldfish* can effect precise conditioned discriminative responses. He can learn to seek the feeding compartment suffused with yellow light and avoid the one flooded with blue light, and this discrimination will be stable and well retained. Goldfish cannot learn some types of complex mazes.

The *white rat* appears as a veritable maze genius and will learn labyrinths with 10 to 15 blind alleys in twenty to thirty trials. Fifty trials may be sufficient to learn a black vs. white discrimination, but the subsequent reversal of the problem will cause considerable difficulty.

Under favorable circumstances the *monkey* will solve a discrimination problem in a single trial and then learn to reverse the discrimination with almost equal readiness. Monkeys exhibit varying degrees of insight in the handling of sticks and boxes as tools.

The *chimpanzee* shows intelligent behavior in the handling of tools. A stick may be used in a wide variety of appropriate ways, as a vaulting pole, a climbing pole, a stick for pushing a banana out of a pipe, or a stick for knocking a banana down from a string. A chimpanzee is intelligent enough to communicate with his fellows. By means of a rudimentary gestural language he may induce another chimpanzee to aid in pulling a food container toward the cage, if the container is too heavy for either animal alone to move. But the chimpanzee never truly attains a spoken language.

The *human being* exhibits an amazing variety of complex performances and his remarkable language ability plays an important role in making these possible. Even as an infant he shows an incomparable variability, plasticity and profusion of verbal utterance. As an adult he usually talks too much.

By comparisons of the capabilities and limitations of the child at different age levels and of the varying performances of animals at different levels of the phyletic series, one can obtain information concerning the nature of the behavioral capacities from the lowest to the highest order. Both child psychology and animal psychology contribute pertinent data to this problem and each field possesses certain advantages. The child is the nearest counterpart of adult man. But the comparative study of animals has the same advantages over child study that it has over the study of adult man himself. Better control can be exerted over the animal's environment. The animal is easier to motivate and to adapt to experimental conditions than the young child. The language of the infant serves often to confuse rather than to aid the analysis of the fundamental nature of many of our complex mental processes.

So far we have discussed the general relationships between development and the increasing complexity of behavior. In the experiments immediately following we shall limit ourselves to a specific aspect of the problem of the nature of the higher mental processes. This is an analysis of the behavior of the rhesus monkey in the solution of problems demanding the use of an abstracted principle. All the problems described utilize a

single basic technique, that of "matching-from-sample." The significance of these problems and their relation to the broader field of complex mental processes are discussed at the end of the section.

SIMPLE MATCHING-FROM-SAMPLE [1]

Benjamin Weinstein

In the typical discrimination experiment the subject is trained always to respond positively to one of two unlike stimuli and negatively to the other. This differentiation is made possible by *repetitively pairing* one of the stimuli with the incentive, a reward or a punishment. The fundamental characteristic of learning of this type lies in the fact that the *same stimulus is always followed by the same incentive* and this unvarying presentation order is maintained throughout the entire training sequence. Thus if blue and yellow stimulus-objects are used, *either* the blue *or* the yellow is correct and *the same stimulus is always correct* no matter how many trials are presented.

Problems of the matching-from-sample type differ sharply from discrimination problems in that in the former *the stimulus paired with the incentive* (*the choice-object stimulus* [2]) *changes from trial to trial;* and both stimuli of this type are rewarded at some time during the course of the training period. Under such circumstances selection of the correct stimulus must depend upon an additional sign. This is supplied by a *sample-object* [2] to which the animal responds first and which is identical from trial to trial with the *correct* choice-object stimulus, i.e., the one presented with the incentive. Thus if blue and yel-

[1] The investigations here presented which use the matching-from-sample technique describe hitherto unpublished studies made possible by grants-in-aid from the Graduate Research Funds of the University of Wisconsin and by assistance from the personnel of WPA Official Project No. 10305. The matching-from-sample technique used is adapted from that already described by B. Weinstein, *J. Comp. Psychol.*, 1941, *41*, 195-213.

[2] See below for a clarification of this term.

low choice-object stimuli are used in a matching-from-sample experiment, *both* the blue *and* the yellow choice-objects are, at one time or another, correct—and incorrect—since the *correct-*

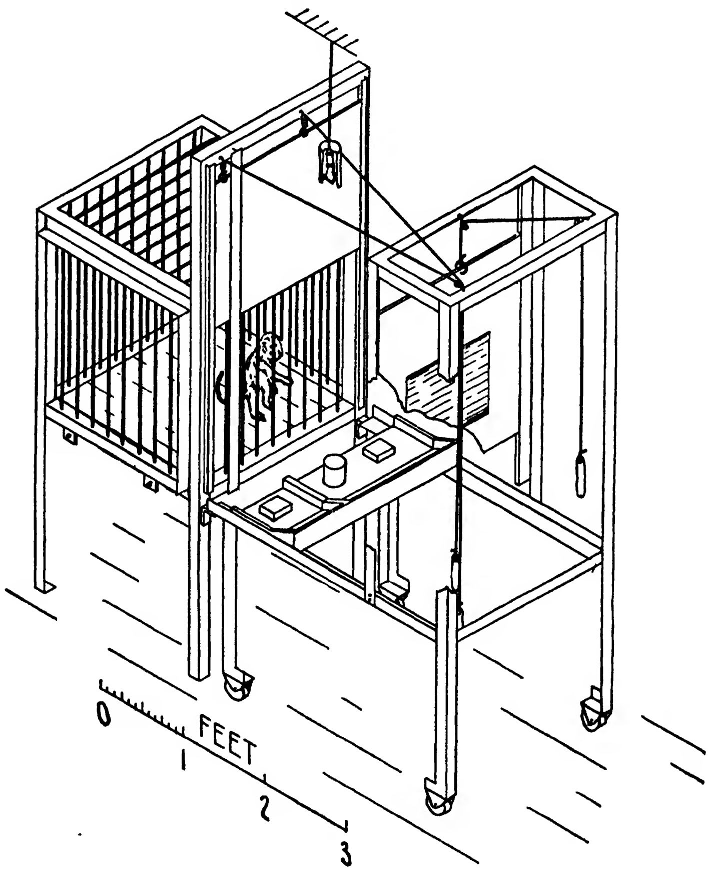


FIG. 25.

ness or incorrectness of a particular choice-object changes from trial to trial, depending upon the nature of the additional sign furnished by the sample-object.

Apparatus. The apparatus used in this experiment is illustrated in Fig. 25 and consists of an experimental cage with a frame in front. The frame is arranged so that a sliding tray may be presented to the subject. Two screens are shown in the draw-

ing. The forward screen is made of metal and is divided into two parts so arranged that either part or the total screen may be raised or lowered. The back screen holds a one-way visual screen, making it possible for the experimenter to observe the monkey while the monkey cannot observe the experimenter. This one-way visual screen was used only in control tests described at the end of the experiment and was not used during the regular testing procedures.

The sliding tray is separated into two divisions, a sample-object-area and a choice-object-area, by a narrow strip of wood and is equipped with three food wells, one in the sample-object-area and two in the choice-object-area.

A large number of objects were used as stimuli and consisted of common household and laboratory articles such as cookie cutters, oil cans, light sockets, trade cartons, and colored and uncolored geometrical wooden and metal forms.

Subjects. Four rhesus monkeys, the sacred monkeys of India, were used as subjects. All these animals had previously undergone an intensive taming procedure and all had had earlier experience on discrimination problems.

Procedure. The procedure was divided into two parts, a preliminary tutoring procedure and a regular training procedure.

In the tutoring procedure the screens were never lowered and the monkeys were free to observe the entire experimental situation and all manipulations made by the experimenter. Under these conditions the tray with a sample-object and two choice-objects was presented to the subject, just out of his reach. The monkey was oriented to the tray by a piece of banana or grape, highly preferred foods, which was held above it. As the subject looked on, the incentive was placed within one of the food wells in the choice-objects-area and covered with the positive choice-object which was the duplicate of the sample-object of the current trial. Simultaneously, the negative choice-object was placed over the empty food well. The animal was next oriented toward the sample with another piece of banana or grape which was placed into the food well and cov-

ered with the sample-object. The sliding tray was then pushed within reach of the monkey who was trained to push aside the sample-object, secure the food, shift over to his left, displace the identical choice-object and obtain the second reward.

When the subject had learned to begin each trial by pushing aside the sample-object, a piece of apple, a non-preferred food, was placed beneath the sample-object instead of a piece of banana or grape. The preferred food was obtained now by the monkey only after choosing the correct choice-object.

As soon as the subject made 23 correct choices in the 25 consecutive trials which constituted a day's run, the preliminary tutoring was concluded and the regular training begun.

In this procedure the forward screen separated the monkey from the experimenter and from the sliding tray during the time that the food and the stimulus-objects were being placed into position. The screen was then raised and the tray slowly pushed forward. The animal first responded to the sample and then chose either the positive or the negative choice-object. If an error was made, i.e., if the non-matching choice-object was displaced, the tray was retracted and the unobtained reward was shown to the monkey by lifting the positive choice-object. The animal was never permitted to obtain the reward after an error.

In all the problems a predetermined trial series was used in which the positive choice-object and its position were varied in an irregular sequence. Neither the same positive choice-object or the same positive position were ever presented more than three times in succession. The same choice-object and the same position frequently changed from positive to negative in successive trials.

Results: Matching-from-sample with two pairs of choice-objects. All four subjects learned to match-from-sample, using two pairs of stimulus-objects, small tobacco cans and small oil cans. The number of preliminary tutoring trials and regular training trials taken by these subjects to reach a criterion of 23 correct responses in 25 trials and the number of correct choices in the last 25 trials are given in Table 3.

TABLE 3

NUMBER OF TRAINING TRIALS AND NUMBER OF CORRECT CHOICES IN THE LAST 25 TRIALS IN THE SOLUTION OF THE MATCHING-FROM-SAMPLE PROBLEM

Subject's number	*Preliminary tutoring trials*	*Regular training trials*	*Total trials to reach criterion*	*Correct choices in last 25 trials (criterion trials)*
50	300	200	500	25
51	550	50	600	24
52	825	325	1150	23
53	700	200	900	23

Transfer of matching-from-sample to two new pairs of choice-objects. This second experiment served as a test of the performance of the subjects in transferring the matching-from-sample behavior to two new pairs of stimulus-objects, blue wooden disks an inch thick and three inches in diameter and small aluminum cookie cutters with green handles. Again, all four subjects readily solved the problem, and the data are summarized in Table 4.

TABLE 4

NUMBER OF TRAINING TRIALS AND NUMBER OF CORRECT CHOICES IN THE FIRST AND LAST 25 TRIALS USING TWO NEW PAIRS OF CHOICE-OBJECTS

Subject's number	*Training trials to reach criterion*	*Correct choices in first 25 trials*	*Correct choices in last 25 trials (criterion trials)*
50	150	19	23
51	200	18	24
52	50	21	24
53	225	19	23

Three factors indicated transfer. No tutoring trials were necessary, the total number of trials taken to reach the criterion was reduced and the number of correct choices in the first 25 trials exceeded chance.

Transfer of matching-from-sample to 25 pairs of choice-objects. After the subjects had demonstrated the ability to transfer the principle of matching-from-sample to two new pairs of objects, the further extension of the transfer process

to 25 new pairs of stimulus-objects was attempted. Each pair was presented only once in a day's run. All subjects effected this complex transfer as is shown by the data of Table 5.

TABLE 5

NUMBER OF TRAINING TRIALS AND NUMBER OF CORRECT CHOICES IN THE LAST 25 TRIALS USING 25 NEW PAIRS OF CHOICE-OBJECTS

Subject's number	*Training trials*	*Correct choices in first 25 trials*	*Correct choices in last 25 trials*
50	75	22	24
51	75	20	25
52	75	17	24
53	50	22	24

After the completion of this training all 4 subjects were tested on 25 additional pairs of stimulus-objects. In the 25 trials run, subject 50 made one error, subject 51 three errors, subject 52 six errors, and subject 53 one error.

Finally, control tests were carried out to test the possibility that the animals were responding on the basis of secondary cues. In these tests the forward screen was lowered and the sliding tray with appropriately loaded food wells was pushed into position close behind it. The one-way visual screen was then lowered, the forward screen raised and the subject allowed to make his choice. The original 25 pairs of stimulus-objects were used and all four subjects attained the criterion of 23 correct responses in 25 successive trials.

Summary and conclusion. The above data show clearly that rhesus monkeys can be trained to respond to the principle of matching-from-sample, and that once this principle is acquired it can be readily transferred to a large number of new sets of stimulus-objects.

DISCRIMINATIVE MATCHING-FROM-SAMPLE

Benjamin Weinstein and H. F. Harlow

The discriminative matching-from-sample problem represents a complication over the simple matching-from-sample

problem in that two sample-objects and five choice-objects are presented to the monkey on the sliding tray. In any one trial, one and only one of the sample-objects can be moved and this is designated as the correct sample-object.

The problem facing the monkey in this particular test is to determine by exploration the correct sample-object and then to select the choice-object which matches the movable sample-object.

Subject and apparatus. A single subject was used in the following experiment. This subject had previously been trained to solve a matching-from-sample problem, with one sample-object and four choice-objects, and he had also been trained to match even when no food was placed under the sample-object.

The apparatus was the same as that used before save for the fact that the sliding tray contained two food wells in the sample-objects-area and five food wells in the choice-objects-area. Furthermore, the food wells in the sample-objects-area were made of machined brass cups so designed as to form a collar that would hold immobile a socket attached to a stimulus-object.

Two sample-objects and five choice-objects were always presented. Each of the sample-objects had a corresponding choice-object, and each of the other three choice-objects was unique. A total of five matching stimulus pairs were used. The general nature of the apparatus is illustrated in Figs. 26*a* and 26*b*.

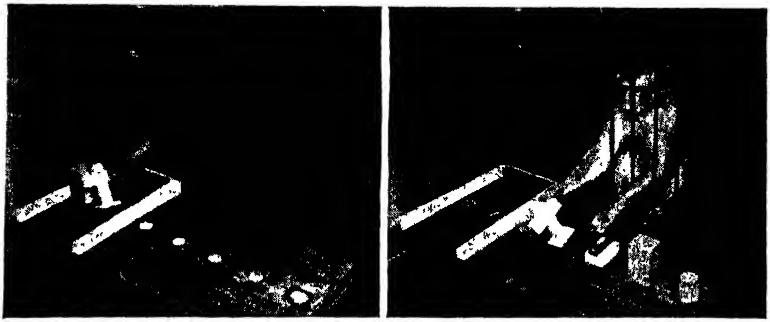
Throughout the experiment food was found only under the correct choice-object.

Procedure. During the first day's run of 25 trials the experimenter assisted the animal in his orientation to the problem. On several trials the monkey repeatedly tried to push aside the incorrect sample-object but made no effort to respond to the correct sample-object. In these cases the experimenter lifted the correct sample-object from the tray, allowed the monkey to touch the correct sample-object and replaced it on the board. The trial would then be repeated. On other trials the subject selected an incorrect choice-object but made no effort to vary

his choice until the correct choice-object was found. In these cases the experimenter lifted the correct choice-object from the board and exposed the food which lay beneath it. The trial was then repeated.

Persistent incorrect responses to both the wrong sample and its corresponding choice-object were discouraged by verbal reproof throughout the entire experiment.

After the first day's run the subject consistently responded



a. *b.*

FIG. 26. Discriminative delayed matching-from-sample. *a.* Selection of sample. *b.* Choice after delay interval.

to the correct sample-object, either immediately or after first trying to displace the incorrect sample-object.

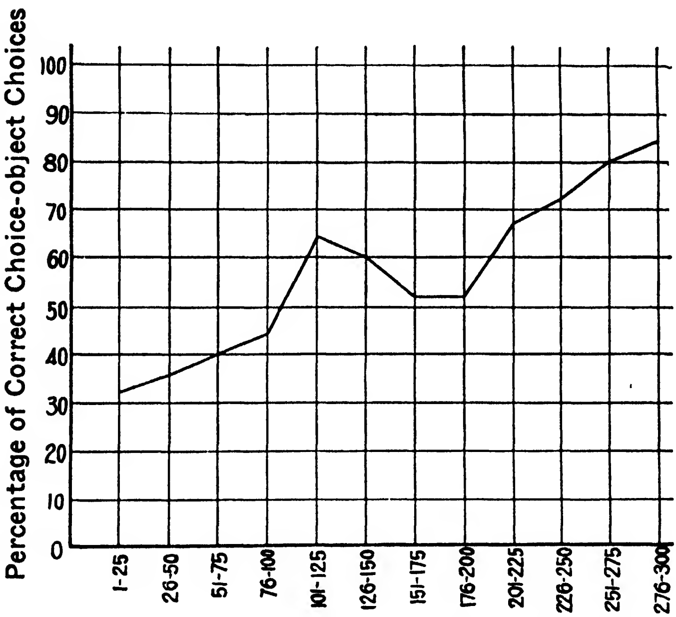
The correct sample-object and the position of the correct sample-object was changed in an irregular but predetermined manner and the arrangement of the choice-objects was such that all choice-objects appeared with equal frequency in all positions for each sample-object configuration.

Results. Before making the correct choice-object selection, the subject reacted to the sample-objects with one of the three following types of response: *Type I, choosing the correct sample-object; Type II, choosing the incorrect sample-object;* and *Type III, attempting first to move the incorrect sample-object and, when this failed, going over to the correct sample-object and pushing it aside.*

Since the Type I response permitted the problem to be

solved as a simple matching-from-sample situation (if the second sample was ignored), little value was attached to it as an indicator of the discriminative matching-from-sample solution. Type II was clear-cut as a correct response, but error responses

PERCENTAGE OF CORRECT CHOICE-OBJECT CHOICES
MADE TO TYPE III SAMPLE MANIPULATION



Type III Sample Manipulation in Groups of 25 Trials

FIG. 27.

utilizing the same principle could not be differentiated from attempted direct matching of the incorrect sample-object. Type III responses always involved discriminative matching since both sample-objects were manipulated. Both correct and incorrect choice responses were easily observed and recorded for responses of this type. For these reasons the correct solution of Type III responses was taken as the primary indicator of the discriminative matching-from-sample solution.

Many different types of error responses occurred but they need not concern us here.

Throughout the experiment the subject initially chose the correct sample-object in about half the trials. In these cases he immediately matched the selected stimulus and there were few recorded errors in over 300 trials. At the beginning of the experiment the subject made many perseverative wrong-sample responses in which he would try from 2 to 6 successive times to displace and match the incorrect sample without recourse to the correct sample. These responses dropped out almost completely after the first 125 trials. With successive trials, a steady reduction of matching errors following Type III sample manipulation was observed. In Fig. 27 is presented a curve showing the increases in percentage of correct choice-object choices made following sample-object manipulation of Type III. Since there were five choice-objects, a correct choice should have been made only 20 per cent of the time by chance.

Summary and conclusions. The discriminative matching-from-sample problem, involving the determination by exploration of the correct sample-object and the subsequent matching of this stimulus with the appropriate one of five choice-objects, was solved by a rhesus monkey.

Discriminative Delayed Matching-from-Sample

Benjamin Weinstein

The delayed reaction experiment remains the classical experiment in comparative psychology designed to show the existence of some "representative factor" enabling the animal to make a discrimination in a situation in which no external differential cue has existed during an interval of delay. If bodily orientation is excluded, the differential cue at the moment of choice must be some central process, imaginal or symbolic in type.

Most delayed reaction studies pertain to spatial delayed reactions in which the animal need only remember the *position*

of the correct stimulus-object as differentiated from the incorrect. Yerkes and Nissen, however, have recently demonstrated that the chimpanzee is capable of responding to non-spatial delayed reactions. A non-spatial delayed reaction is a delayed reaction so planned that the subject cannot solve the problem unless he remembers some *quality* of the correct stimulus-object other than its position. The representative factor must provide a signal of *what* the object is, not *where* the object is.

In the following experiment the matching-from-sample technique was used to test the performance of a rhesus monkey in the solution of problems involving non-spatial delayed responses.

Apparatus. The apparatus is the same as that used in the discriminative matching-from-sample tests and is illustrated in Figs. 26*a* and 26*b*.

Subject. A single subject was used in this experiment. This animal had already solved the discriminative matching-from-sample test and also some delayed reaction matching-from-sample tests of 15-seconds duration in which the single sample-object and two choice-objects situation was used.

Method. Before each trial of the discriminative delayed matching-from-sample problem was run, the forward screen was lowered so that it was impossible for the monkey to see either the experimenter or the test situation. Two sample-objects, a correct sample-object which was movable and an incorrect sample-object which was fixed, were then placed in position on the sliding tray which was pushed forward close to the experimental cage. The narrow section of the opaque metal screen was raised and the animal was allowed five seconds to discover and examine the correct sample-object (see Fig. 26*a*). The narrow section of the opaque screen was next lowered and both sample-objects removed. The delay interval began with the closing of this screen section, and during this interval the five choice-objects were placed in position and food was put within the food-well of the correct choice-object. At the expiration of the delay interval the wide section of the screen was raised and the monkey was allowed to make his choice

(see Fig. 26B). At the time of the subject's choice the experimenter was hidden by the one-way-visual screen and had no contact with the sliding panel. This situation afforded a check against any secondary cues.

Even though the subject was familiar with both the delayed matching-from-sample technique and with the discriminative matching-from-sample technique, approximately 200 preliminary trials at delay intervals ranging from 0 to 10 seconds were run. At the end of this time the regular test trials were begun. Five 0-second and ten 15-second delayed reaction trials were run for 15 days. At the end of these trials additional data was obtained at 0-second, 15-second, 30-second and 60-second delay intervals. For each of 20 additional days five delayed reaction trials were given at each of the four delay intervals, the shorter delays being given first and the longer delays being given last.

Results. The results (exclusive of those for the 0-second delay intervals) are presented in Table 6. For each delay interval the data is treated in groups of 50 trials. The total number of responses and the percentage of correct responses are tabulated in terms of Type I and Type III sample manipulation, described on page 186. Type I responses are those in which the monkey chose the correct sample-object first; Type III responses those in which the subject chose the incorrect sample-object first and then shifted to the correct sample-object. A chance score is 20 per cent correct, since 5 choice-objects were used. Any score over 50 per cent correct for 25 trials is more than 3 S.D. beyond chance, and is, therefore, highly reliable.

Comparison of the first 100 trials with the later trials at a delay interval of 15 seconds indicates that the monkey improved with practice. Comparison of the last 100 trials at 15-seconds delay with the 100 trials at delay intervals of 30 and 60 seconds shows that the performance of the subject decreases in efficiency with increasing length of delay. Even at a delay interval of 60 seconds, however, the responses of the subject

are reliably beyond chance for both Type I and Type III sample manipulation.

TABLE 6

PERCENTAGE OF DISCRIMINATIVE-DELAYED MATCHING TRIALS CORRECT AT THREE INTERVALS OF DELAY

	15-sec. delay				*30-sec. delay*				*60-sec. delay*			
	Type I		*Type III*		*Type I*		*Type III*		*Type I*		*Type III*	
Trial number	*No*	*%C*	*No*	*%C*	*No*	*%C*	*No*	*%C*	*No*	*%C*	*No*	*%C*
1- 50	24	67	26	42	22	91	28	68	22	42	28	50
51-100	20	80	30	50	28	79	22	73	24	54	26	65
101-150	22	82	28	75								
151-200	19	84	31	90								
201-250	26	90	24	84								

No = Number of responses.
%C = Percentage of correct responses.

Summary and conclusions. The data of this experiment show clearly that the rhesus monkey is capable of making non-spatial delayed responses after intervals of 30 to 60 seconds. Even when the experiment is complicated by forcing the subject to first effect a discriminative response, non-spatial delayed responses of at least 30 seconds are successfully accomplished.

Since the above problem cannot be solved by bodily orientation, one is forced to assume that some imaginal or ideational brain process operates during the delay interval and makes possible a correct solution when the choice-objects are presented. In the differential delayed responses of Type III it would appear that the imaginal process is not merely the representation of an object, but is a discrimination between objects.

ALTERNATION OF THE MATCHING-FROM-SAMPLE AND THE NON-MATCHING-FROM-SAMPLE PROBLEMS ACCOMPANIED BY APPROPRIATE CUES

H. F. Harlow

None of the problems previously described can be solved by consistently responding in a positive or a negative manner to any particular stimulus or stimulus pattern, but a *principle*—response to the *sameness* between the sample and the correct

choice-object—must be employed if solution of the problem is to be attained.

The following experiments introduce an important complication, in that here the monkey is required to utilize *two antagonistic principles:* (1) the selection of a choice-object which matches a sample-object and (2) the selection of a choice-object which does not match a sample-object, when these two principles are introduced in a random manner.

Two cues served as indicators of the antagonistic principles. In the matching-from-sample situation the stimulus-objects were placed on a *black sliding panel* and a piece of non-preferred food was placed under the sample-object. In the non-matching-from-sample situation the same stimulus-objects were placed on a *yellow sliding panel* and no food was placed under the sample-object.

Apparatus. The apparatus was the same as that previously described save that a second sliding tray was used. This sliding tray was painted a bright yellow and was not divided by any partition separating the sample-objects-area from the choice-objects-area.

Two pairs of stimuli, green crosses 2½ inches high and 1½ inches wide, mounted on two-inch square bases, and two white electric push buttons with copper bases, a little less than 2 inches in diameter, were used in the original training.

Subjects. The subjects in this investigation were two rhesus monkeys.

Procedure. Since the subjects used had already been trained to do matching-from-sample, it was first necessary to teach them the opposing non-matching-from-sample principle. To effect this, preliminary non-matching-from-sample training was carried out using the yellow sliding panel. The general method used was that employed in the preliminary matching-from-sample tutoring trials save for the fact that now food was placed only under the choice-object which did not match the sample. The animals' response was scored as correct if he either pushed aside the sample-object and then the non-matching choice-object, or if he pushed aside only the non-matching

choice-object. Twenty-five trials a day of non-matching training was given.

Each day's non-matching trial series was always followed by matching-from-sample trials until 5 successive correct matching-from-sample responses were made. The black sliding panel was, of course, used for the matching-from-sample trials.

As soon as the subject attained a criterion of 90 per cent correct in the non-matching trials for two successive days in this preliminary tutoring, regular non-matching training trials were initiated. In these trials the opaque metal screen was lowered in front of the sliding tray, while the food-well beneath the correct stimulus, the non-matching choice-object, was loaded with a preferred food. Twenty-five non-matching trials were again run each day until the subject had attained a criterion of 90 per cent correct for 50 successive trials. Five successive successful matching-from-sample trials were always run at the conclusion of each of the non-matching series.

When the criterion for the solution of the non-matching-from-sample problem had been attained, the problem of alternation of matching-from-sample and non-matching-from-sample was begun. This training was divided into two stages.

In stage I, five trials of non-matching-from-sample using the yellow sliding panel were alternated with five trials of matching-from-sample using the black panel. Fifty test trials were run each day, 25 matching-from-sample trials and 25 non-matching-from-sample trials. The first trial of each series of five was given especial consideration as a reversal trial, since on these trials the subject had to utilize a principle antagonistic to the principle which was correct on the previous trial if he were to successfully solve the task at hand.

Stage II utilized the method of presenting matching-from-sample trials and non-matching-from-sample trials in four irregular but predetermined series.

On both stages I and II of the alternation of principle problem the monkeys were trained to a criterion of 90 per cent correct in 50 successive trials for all situations: the matching, the non-matching and the reversal situations.

Results. The number of trials taken by the two animals to attain a criterion of 90 per cent correct responses both in the preliminary tutoring and in the standard training for non-matching-from-sample is presented in Table 7. The number of correct responses in the last 50 trials is also given.

TABLE 7

NUMBER OF TRAINING TRIALS AND NUMBER OF CORRECT CHOICES IN THE LAST 50 TRIALS OF TRAINING FOR NON-MATCHING-FROM-SAMPLE

	Trials to attain criterion		*No. responses correct in last 50 trials*	
Subject's number	*Preliminary tutoring*	*Standard training*	*Preliminary tutoring*	*Standard training*
50	125	200	45	47
51	125	175	45	47

In Table 8 are presented the data for stage I of the problem involving the alternation of matching-from-sample with non-matching-from-sample. These data were obtained by running the matching and non-matching situations in alternate groups of five trials. This table lists the number of trials taken before the subjects attained a criterion of 90 per cent correct on all of three indices in a single group of 50 trials. These three indices were: (1) correct matching trials, (2) correct non-matching trials, and (3) correct trials involving a shift from one principle to its opposite. The table also lists the number of correct trials for each of these indices in the last day's run. Twenty-five correct matching and non-matching trials and 10 correct shift trials would represent a perfect score.

TABLE 8

RESPONSES TO MATCHING-FROM-SAMPLE AND NON-MATCHING-FROM-SAMPLE WITH ALTERNATIONS OF THE OPPOSING TEST SITUATIONS EVERY FIVE TRIALS

	Training trials	*No. correct responses in last day's run*		
Subject's number	*to attain 3 criteria*	*Matching trials*	*Non-matching trials*	*Shift trials*
50	150	23	23	9
51	200	23	25	10

Comparable data for stage II of the alternation of matching-from-sample and non-matching-from-sample problems are presented in Table 9. This phase of the general problem was run with the antagonistic situations presented in a predetermined irregular order. It should be noted here that there are 25 instead of 10 shift trials in every day's run.

TABLE 9

RESPONSES TO MATCHING-FROM-SAMPLE AND NON-MATCHING-FROM-SAMPLE WITH ALTERNATIONS OF THE OPPOSING TEST SITUATIONS IN AN IRREGULAR ORDER

Subjects	*Training trials to attain 3 criteria*	*No. correct responses in last day's run* Matching trials	Non-matching trials	Shift trials
50	100	24	23	23
51	100	23	25	24

Summary and conclusions. In this experiment the differential cues furnished by the color of the sliding tray and the reward or lack of reward beneath the sample object call forth a differential principle making possible the problem solution. Thus the rhesus monkey is not only capable of solving a problem involving a principle such as that of matching or non-matching but can alternately use both of these antithetical principles if some differential cue is available to elicit the appropriate principle from trial to trial.

GENERAL DISCUSSION OF SECTION

The relation between the matching-from-sample problems and the problems of simple discrimination can best be seen by comparison of the response to the correct choice-object in the matching-from-sample situation with the response to the correct position or stimulus-object in the discrimination experiment.

In a spatial discrimination, in which the subject learns always to choose the left or the right stimulus-object, a constant relationship between position and incentive is maintained throughout a trial series. Response to the left or to the right position, depending upon the one designated correct, is always followed by a reward. According to Tolman the position be-

comes a *sign* for the incentive. By sign Tolman refers to the fact that a response to a particular stimulus, in this case that of position, arouses in the subject an attitude of expectancy or anticipation toward the re-enforcing agent, the food.

Discriminations of object-quality between unlike objects differing in one or many characteristics (size, hue, saturation, brightness, etc.) are similar to spatial discriminations in that a constant relationship is maintained throughout a trial series between the object-quality of the correct object and the incentive. Responses to the correct object are always rewarded. Here the total qualitative pattern of the object, as distinguished from its position, becomes a sign for the incentive.

In the response to the choice-object in the matching-from-sample problem neither position nor object-quality can act as a sign for the incentive since both the position and the object designated as correct vary in an irregular order from trial to trial. To solve the problem the animal must first respond to the similarity between sample and correct choice-object.

Here the sample object serves as a sign to indicate which choice-object is a sign for an incentive. This might be called *sign-of-a-sign behavior* as differentiated from the sign-behavior involved in the simple spatial or object-quality discrimination.

The alternation of the matching and the non-matching situation introduces a further complication in that the sample-object per se no longer acts as a sign to designate correctness between the choice-objects. Here either the color of the sliding tray or the presence of food under the sample-object acts as a sign which further differentiates the value of the sample-object as a cue. Expanding Tolman's terminology this might be called *sign-of-a-sign-of-a-sign behavior.*

Discrimination problems as well as matching-from-sample problems can only be solved if the subject can form an appropriate discriminative set. Differences in complexity in these two basic situations are to be referred to the more complicated stimulus conditions which are involved in the solution of the matching-from-sample situation as compared to the discrimination problem.

CHAPTER 13

The Inheritance of Psychological Characteristics

MOST studies concerning the inheritance of psychological traits have dealt with such categories as intelligence, criminality, insanity and feeble-mindedness. The behavior which characterizes these traits is highly variable and complex; social and cultural factors are difficult or impossible to control; and basic breeding controls, which the geneticists have developed in animal experimentation, are impracticable with human subjects. As a result we possess a vast literature of conflicting claims, with little definitive knowledge of the actual mechanisms involved in psychological inheritance.

Such limitations in this important field of human investigation clearly indicate the need for rigidly controlled scientific studies in which the inheritance of psychological abilities is studied in the same manner that the inheritance of physical and physiological characteristics are investigated.

We have already pointed out that adult human beings are poor subjects for such investigations since they live to excessive ages, mature slowly, refuse to lead rigidly controlled lives, subject themselves to varied and unpredictable environments, and often breed in response to passing fancies instead of following sound genetic procedures.

If we are to have a science concerning the inheritance of psychological characteristics it must be based upon prolonged and carefully controlled studies in which circumscribed and closely defined abilities are investigated in sub-human animals.

Relatively few investigations of this type have as yet even been initiated, because of the great amount of time and effort that they necessitate. Such studies as we do have, however,

indicate the major lines of approach to be used and show clearly that valuable information can be obtained.

The two studies on the inheritance of psychological traits described in this section have been chosen because they are investigations within different broad psychological categories, the general fields of temperament and of learning and intelligence.

Wildness and Savageness in Rats of Different Strains

C. P. Stone[1]

The albino rat, widely used in many different fields of scientific investigation, is known to have been domesticated for a hundred years. Selective breeding, either deliberate or unintentional, has produced an extremely tame and easily handled animal. Since the albino rat is merely a domesticated breed of the wild Norway rat, it is entirely feasible to compare under standardized laboratory conditions the albino rat with the wild rat or with any mixture of the two breeds. Thus pertinent data can be obtained concerning the influence of hereditary factors produced by matings between strains upon any desired traits.

Wildness and savageness was the particular trait chosen for investigation in the study to be described.

Subjects. The subjects used in this experiment were 17 pure albino rats from the Wistar colony, 10 wild animals trapped in a basement near a large hospital within a residential district of Chicago, 70 first-generation crosses between the trapped animals and the tame albinos, and 30 crosses between the half-breeds and the albino rats. These groups will subsequently be designated by the following abbreviations:

A = albino; QB = quarter-breeds, crosses between albinos and half-breeds

[1] "Wildness and Savageness in Rats of Different Strains." In Lashley, K. S. (Ed.), *Studies in the Dynamics of Behavior*. Chicago, University of Chicago Press, 1932, pp. 3-55.

HB = half-breeds; FW = full-wilds

Apparatus and method. Two distinct techniques were used in the determination of the wildness and savageness of the four groups of rats, one involving a rating of observed behavior and the other an objective test.

The rating method entailed the observation of the behavior of the rat in two standard situations and the rating of the subjects on a three-step scale for 11 different traits in each, as indicated below.

Situation No. 1. Capturing and removing the rat from the home cage; holding him in one hand for 1 to 2 minutes while he was inspected for ear marks, tenseness of body, urination, etc.

Situation No. 2. Placing the animal, with others of his group, on the floor of a small inclosure, approximately 5 x 6 feet; allowing the subject to remain free and without disturbance for approximately 30 minutes; recapturing subject, placing it on a screen of hardware cloth and noting general behavior and attempts at escape when momentarily released and when restrained. These observations to last from 1 to 2 minutes.

The items for which each strain of animals was rated were:

1. resistance to catching in the home cage
2. tenseness of muscles of the body
3. jumping, clawing, squirming, wriggling
4. squealing when caught or held
5. urination
6. defecation
7. biting
8. threatened biting
9. laying back ears
10. gnashing of teeth; chattering
11. hissing noise

With the rating scale used, a score of 0 stood for the absence or minimal expression of a trait. The upper limit, or value 3 for a given item, was set by the behavior of wild brown rats

tested soon after they had been trapped and confined in the colony room.

Since wildness scores of the type described are markedly influenced by previous handling every effort was made to keep this factor constant. Ratings on all groups save for the full-wild animals were made between three and four months of age and no preliminary handling had occurred save for the removal and transfer of the young from the mother at weaning time. Feeding of the rats and cleaning of the cages were effected without disturbing or touching the animals.

The *objective test situation method* provided a measure of the relationship between the hiding tendency and hunger. The apparatus consisted of a starting box and a goal box connected by a U-shaped stovepipe-tunnel. An exit door in the starting box and an entrance door in the food box prevented retracings after these compartments had been left or entered, respectively. A small pedal was placed in front of the food receptacle and this pedal operated a buzzer through an electrical contact which signaled to the experimenter the completion of the trial. One trial a day was given and the subject was allowed 10 minutes to complete the circuit. In case of failure to complete the run, no food was given for that day. Strong initial motivation was provided by depriving the subjects of food for 48 hours before the beginning of the experiment.

Results: Rating of observed behavior. The comparative influence of ten successive days of handling upon the reduction of wildness and savageness in the four groups is shown in Fig. 28. All groups become tamer with handling, but the wild rats and half-breeds never become as docile as the quarter-breeds and albinos. Indeed, at the end of the ten days of handling the wild rats were not as tame as the albino rats on the first day of observation.

The mean ratings received by each group on days 1, 5 and 10 for both situations are presented in Table 10, and the data showing the significance of the differences of these means are given in Table 11. Inspection of this latter table shows that al-

most all group differences are highly significant save for the differences between the albinos and the quarter-breeds.

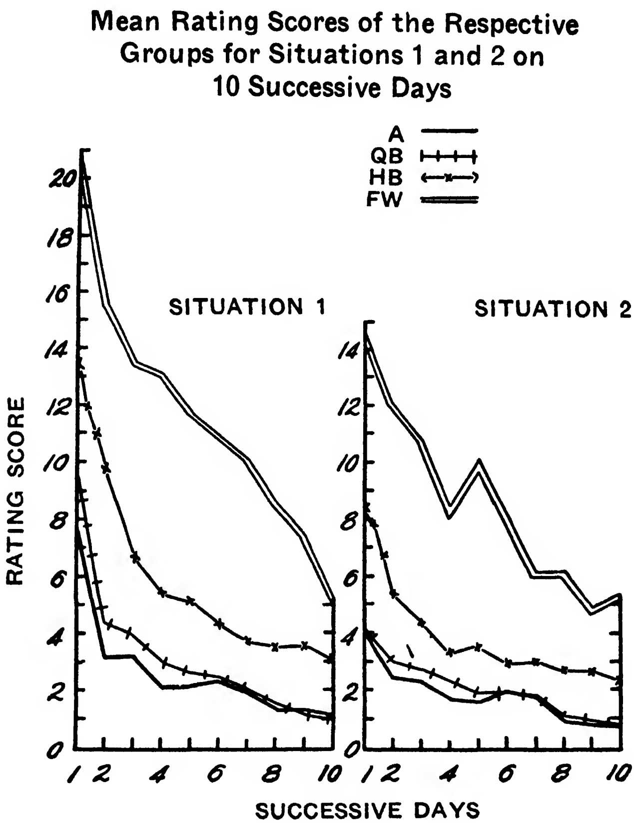


FIG. 28.

TABLE 10

MEAN RATINGS RECEIVED BY EACH GROUP ON DAYS 1, 5 AND 10 FOR BOTH SITUATIONS

	Situation 1			*Situation 2*		
Group	*Day 1*	*Day 5*	*Day 10*	*Day 1*	*Day 5*	*Day 10*
A	7.5 ± .69 *	2.1 ± .18	1.1 ± .12	3.9 ± .27	1.6 ± .17	0.7 ± .10
QB ..	9.2 ± .43	2.6 ± .18	1.0 ± .12	4.0 ± .17	1.9 ± .17	0.7 ± .09
HB ..	13.5 ± .27	5.0 ± .16	3.0 ± .12	8.3 ± .22	3.5 ± .14	2.3 ± .09
FW ..	20.3 ± .71	11.6 ± .77	5.3 ± .85	14.2 ± .96	10.0 ± .69	5.3 ± .82

* Standard deviations.

TABLE 11

CRITICAL RATIOS OF THE DIFFERENCES BETWEEN MEANS OF THE RATINGS FOR DAYS 1, 5 AND 10 FOR BOTH SITUATIONS

	Situation 1			*Situation 2*		
Groups	*Day 1*	*Day 5*	*Day 10*	*Day 1*	*Day 5*	*Day 10*
A vs. QB	2.17	1.80	0.17	0.21	1.33	0.46
A vs HB	8.15	11.99	11 53	12.43	8.45	12.69
A vs. FW	12.96	12.08	4.92	10.27	11.83	5.58
QB vs. HB	8.37	10.20	11.70	15.28	7.00	12.23
QB vs. FW	13.33	11.41	4.95	10.46	11.38	5.53
HB vs. FW	8.95	8.32	2.64	6.04	9.34	3.63

As a control for the effect of early conditioning of the young by the mothers, some full-wild young were reared by albino mothers from birth and some albinos were similarly placed with a full-wild mother and young. Under these circumstances the full-wild young when tested yielded rating scores which approximated the mean of their genetic strain rather than the mean of their foster-mother or foster litter-mates. The albinos were no more wild or savage than other albinos reared by their own mothers.

Objective test situation. The percentage of animals in each of the four groups that failed to run the labyrinth and enter the food box within 10 minutes for each of the 10 trial days is shown in Fig. 29. For trial 1 approximately 80 per cent of the wild rats, 65 per cent of the half-breeds, 35 per cent of the quarter-breeds, and none of the albinos failed to enter the food box. Since the hunger drive tends to gain dominance over the hiding tendency (hiding in the maze instead of running to the food box), the great initial differences decrease with practice. The four groups of animals, however, maintain their relative positions throughout the ten days. Actually the group differences in the later trials are probably greater than the curves indicate; since if a rat fails to solve the problem on any one day he receives no food and this will lead to increased organic tensions on the following day.

Summary and conclusion. Albino rats, wild rats, half-breeds and quarter-breeds (one-fourth wild) were tested for wildness and savageness by means of three independent meas-

ures. All three measures give substantially similar data and indicate that wildness and savageness increase as one progresses from albinos to quarter-breeds, to half-breeds, to full-

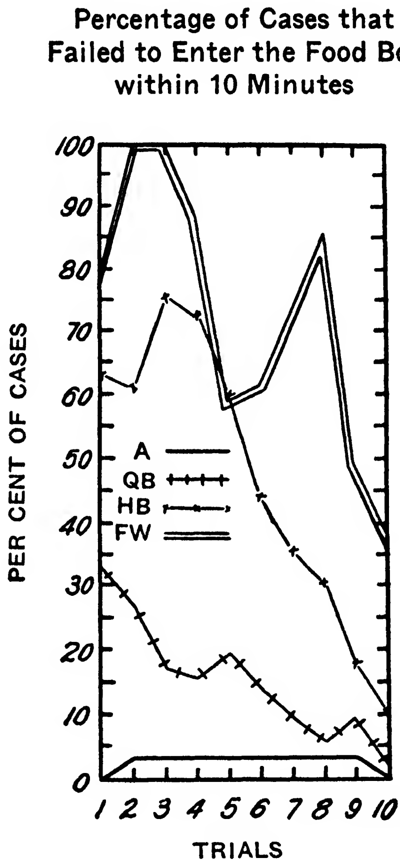


FIG. 29.

wilds. Since the test conditions were kept highly constant, one may conclude that innate temperamental differences exist between these four strains of animals. Controls for the effect of early environmental influences indicate that genetic factors are more important than maternal association throughout the suckling period.

Genetic Differences in Maze-learning Ability in Rats

R. C. Tryon [2]

The basic assumption behind any eugenics program for man is that psychological traits could be improved by selective breeding. Since it is difficult or impossible to effect all needed controls with human beings as subjects, an experimental genetics focused on animal behavior can supply useful information concerning any such problem.

None of the behavorial tests yet devised for sub-human animals is better adapted to a psychological study of inheritance than is the maze, since well-constructed mazes are highly reliable and are simple enough in operation to permit the testing of a relatively large number of animals.

The purpose of the following experiment was to develop under rigid environmental control a maze-bright and a maze-dull strain of rats.

Subjects and apparatus. Eighteen generations of rats served as subjects in this experiment with the breeding schedule so arranged that in each generation the rats showing the greatest maze ability were mated together, and the rats demonstrating the poorest maze ability were likewise mated together. Once the separation of strains was started there were no matings between strains, even though on the basis of performance an occasional animal seemed to belong with the opposite strain. A multiple T type maze so devised that an automatic mechanical device delivered the rats into the maze without handling was used, and an electrical recorder scored each rat's run. All animals were run for 19 trials and the total number of blind alley entrances constituted the final score.

Results. The results of this study in selective breeding of rats for maze-brightness and maze-dullness are outlined in Fig. 29. Data concerning the total blind alley entrances for the original unselected sample of 142 animals and for the bright

[2] "Genetic Differences in Maze-learning Ability in Rats." *Intelligence: Its Nature and Nurture.* 39th Yearbook, National Society for the Study of Education, 1940, pp. 111-119.

and dull progeny of the first, third, fifth and seventh generation are presented in terms of the percentage of each group falling within a particular error score interval. The number of

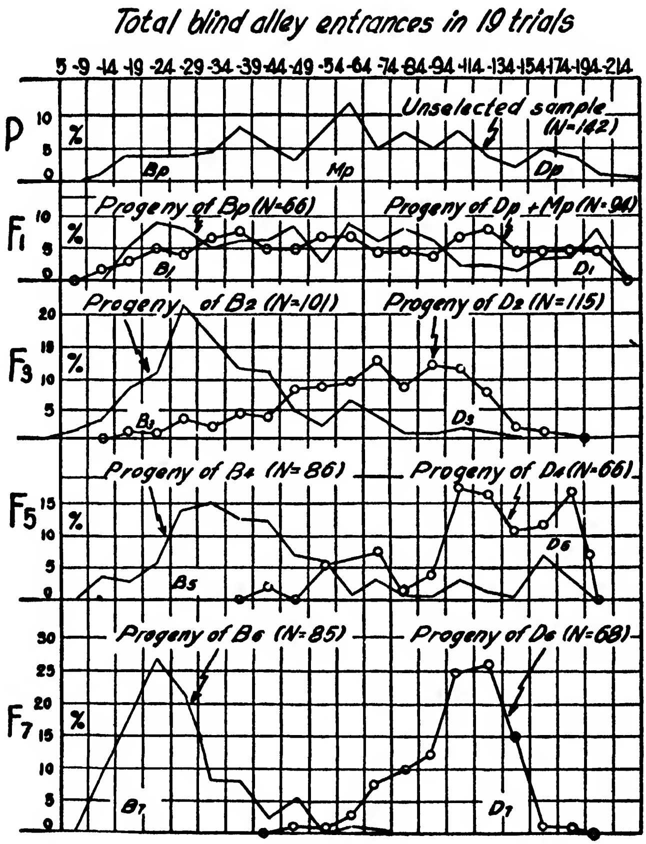


FIG. 30.

rats in each group for these four generations is given in Fig. 30.

The original group is characterized by extreme dispersion of individuals and a unimodal distribution of the total blind alley entrances. Little change appears in the F_1 generation, but by the third generation the range of error scores for both the bright and the dull groups has begun to constrict and the modal points of the two distributions have begun to separate. These

same changes continue throughout the F_5 generation, and by the time generation F_7 is reached, the two distributions have become almost completely separate. Further data are not presented since from the F_8 generation on selective breeding produced negligible effects in the further separation of the total error score curves.

Summary and conclusion. Selective breeding of an original group of 142 rats on the basis of brightness and dullness in learning a complex T maze produced almost complete separation of the total error score distributions made by the members of the two groups of the F_7 generation. Little or no additional separation between the two groups occurred thereafter.

These data prove that it is possible in a few generations to separate rats into two genetic groups in regard to a psychological trait, that of maze performance.

GENERAL DISCUSSION

Both of the studies just described show clearly that valuable information concerning the inheritance of psychological traits can be obtained by rigidly controlled investigations using subhuman animals as subjects.

Each of the two general methods utilized gives satisfactory data. One may either compare pre-existing strains of a single species and their cross-breeds, or one may take an originally unselected group and produce separation by selective breeding.

The data obtained make it seem likely that innate differences may be demonstrated for many other psychological traits that can be measured by reliable and objective test methods. Furthermore, the demonstrated differences dependent upon genetic constitution may appear in a comparatively small number of generations, a fact attesting to the feasibility and immediate value of such researches.

These studies obviously have wide implications concerning the role of heredity as a factor determining human psychological characteristics, but their value is by no means limited to this broad field of theory. The separation of animals into groups on the basis of inherited behavioral tendencies or abili-

ties provides us with ideal subjects for the analysis of the factors, both behavorial and constitutional, that enter into the determination of these differences.

Given groups of subjects of known and constant heredity, it is possible to determine the relative importance of environment in influencing the performance of the various strains. Studies of the influence of such factors as dietary deficiencies, temperature, forced activity, and glandular disturbances upon behavior can be more readily carried out with such subjects.

Tryon has already obtained data on the nature of the sensory abilities and basic behavior patterns found in his maze-bright and maze-dull animals. Investigators at Minnesota have separated rats into active and inactive strains and have studied the relationship between sterility, initial weight, and maze learning and spontaneous activity.

All of these analytical studies of the constitutional factors involved in differences in performance are still in a preliminary stage. The favorable results already obtained, however, suggest that a genetics based on animal behavior studies may not only permit the formulation of basic generalizations concerning the inheritance of psychological traits, but may also provide us with selected subjects making possible more precise studies concerning the nature of the behavorial characteristics and the organic processes involved in the performance of behavior patterns, and the role of environment in influencing these characteristics and processes.

CHAPTER 14

Social Behavior of Primates

IN ALL probability the actual development of man's social organization can never be traced since written records of the early stages of the process are non-existent, and deductions based on human and animal bones, scattered flints and broken pottery can never be confirmed.

Logical deductions, however, concerning the basic mechanisms involved in the development of man's social organization need not be limited to the evidence obtained from the calcium and silica relics which remain the property of the archeologist. The existing primate groups are doubtless variants of man's primordial ancestors and they offer a vast field for primitive sociological studies.

Two general types of investigation appear to be particularly fruitful: field studies of the actual relationships to be found in the existing sub-human primate groups; and detailed experimental studies of those abilities, such as language, imitation and cooperation, which play an important role in the organization of primate social groups. Two researches illustrating these basic fields of study are presented in the following pages.

THE FIELD STUDY

Comparative psychologists are men who frequently know more about apparatus than about the behavior of animals in general. This is no doubt partly occasioned by the stress that has been placed upon the experimental as opposed to the observational method, in spite of the fact that the two are by no means antithetical. Furthermore, many studies in comparative psychology have been dictated by purely practical ends. A pair of breeding rats, a few pounds of McCollum's standard diet, a

little wood and wire, all ingeniously combined, may lead to fame as an experimentalist; whereas a good field study is expensive and time-consuming, and may pay small academic dividends for the patience and ingenuity required.

The values of the field study are many. Knowledge of the organization of animal societies can only be obtained by observations carried out in the field. Even when large numbers of animals of a single species are maintained under favorable conditions in captivity, normal environmental conditions are seriously disrupted with resulting changes in social behavior. Aside from providing knowledge of social organization and social change in animal societies, the field study can offer information concerning basic behavorial adjustments, such as adjustments to food and climate, which provide knowledge essential to the proper maintenance of a stock colony. Furthermore, information concerning the more complex adjustments in the wilds should furnish valuable clues as to the particular potentialities and limitations of various species for experimental investigations.

Behavior and Social Relations of Howling Monkeys

C. R. Carpenter [3]

Barro Colorado Island, lying in Gatun Lake, almost in the center of the Isthmus of Panama, was made a natural reservation about 20 years ago. Trails were cut through the dense forests and over the rugged terrain of the island, and small shelter houses were built at the end of six of these, thus providing excellent facilities for continuous observational work.

The field studies reported here were made during a five-month period in the first half of 1931 and during periods of observation lasting two and one months respectively, approximately a year later.

Subjects: Description. The subjects studied in this investigation were howling monkeys, *Alouatta palliata aequatorialis*.

[3] "A Field Study of the Behavior and Social Relations of Howling Monkeys." *Comp. Psychol. Monog.*, 1934, *10*, No. 2, pp. 1-168.

These new-world monkeys attain a weight of 15-20 pounds and a height of 3-4 feet at maturity. Although arboreal, their locomotion in the trees is "pronograde" in type. They move with the longitudinal axis of the body parallel to the support over which they are passing. The howler monkey does not take long jumps, normal progression being from the terminal branches of one tree to the interlacing terminal branches of another. Their hands, which are poorly adapted for fine prehensile movements, grasp the tree limbs, during locomotion, with the thumb and index finger on one side and the other three digits opposed. The foot is well adapted for climbing, but poorly adapted for walking. The great toe or hallux is opposed to all other toes. The prehensile tail of the howler is in constant use. It functions as a grasping organ during locomotion, an anchor during rest and sleep, and at other times as a manipulatory instrument for driving away insects and for carrying on grooming.

General life habits. The howling monkey's day lasts from sunrise to sunset. Perhaps a quarter of the waking time is spent in feeding. There are two main eating periods, one during mid-morning and one during mid-to-late afternoon. Large quantities of bulky food are consumed, the diet consisting of leaves, fruits and berries. Most of the water consumed comes from the foods eaten, but moisture may be licked from leaves, or collected on the fingers and then licked from them. As the mid-morning feeding period gradually ends, the young ones begin playing and the adults become quiet, and rest or sleep. The mid-day rest period ordinarily lasts from about 11 A.M. to 2 P.M. Progression of the clan along the arboreal routes seldom takes place during this period of the day.

Method. A number of different procedures were used in this field study; these are listed below:

1. *Direct observation with complete concealment of the observer* by means of temporary blinds constructed both in trees and on the ground at places frequently visited by howler clans.

2. *Direct observation from ambush* using the natural concealment afforded by banks, rocks, bushes and tree trunks.

3. *Direct observation following neutral conditioning of the*

animals to the observer, the observer remaining near the monkeys daily for a month or more without threatening or molesting.

4. *Indirect observation: the study of spoor* such as excreta, blood spilled, dead animal remains, and food waste.

5. *Experimentation,* involving the removal of the animals, the introduction of strange animals, or the use of apparatus placed in the wilds. (This method was suggested but actually exploited very little.)

6. *Photography,* including both stills and motion pictures.

7. *Collections,* including animal specimens killed and studied, live animals captured, raised in captivity and observed, and plant specimens of food eaten by the howlers.

Results. The results of this field study are arranged systematically under four main headings: (1) territoriality and nomadism, (2) organization of a howler population, (3) group integration, social relations, and intra-group behavior, and (4) group co-ordination and control.

1. Territoriality and nomadism. Territoriality and nomadism were studied by consecutive daily observations of one clan for a month. During this month and the following four months the clan was located and identified 73 times. A second clan was located and identified 26 times and supplementary data was obtained on 21 other clans.

These data showed there is a strong tendency for a clan to limit its wanderings to a definite, circumscribed area; for example, clan number one was observed only within a territory of 300 acres. Movements within the clan were slow and rather restricted and occurred with reference to food and lodge trees.

The possession of territory was not a static but a dynamic adaptation: territorial ranges were constantly shifting, contracting, or extending, depending upon the season of the year, the available food supply, the formation of new clans and the encroachment of other clans. Some howler groups were relatively isolated, whereas other groups lived concurrently in almost identical ranges.

Howling monkeys cannot be considered nomadic since they

limit their wanderings to a definite area and show marked preference for certain food and lodge trees where they may loiter and rest. The location of food and lodge trees, as well as prominent arboreal routes within the territory of a group, are learned and become a positive influence to the animals.

Analysis reveals four factors which operate to limit the territorial range of groups of howler monkeys: (1) the conditioning of a clan to goals within the area and the acquired facility of reaching these goals over learned pathways, (2) the tendency to react positively to the familiar and to avoid the strange, (3) the tendency to avoid other clans, and (4) the tendency to avoid scrub growth and to show a preference for dense, primary forests.

2. **Organization of howler population.** By careful observation of the various clans it was possible to make a census of the howler monkey population of Barro Colorado Island during both 1932 and 1933. The following data are taken from the 1933 census.

Twenty-eight groups of howler monkeys with a total population of 489 animals were observed. Eighty-two of these were adult males, 192 adult females. Of the females, 98 were mothers, 94 were not. Of the 215 young animals, 98 were infants and 117 were juveniles, i.e., they ranged from about 18 to 36 months in age.

The number of adult males per group ranged from 1 to 5 with a mean of 2.7 and a standard deviation of 1.3. The number of adult females ranged from 3 to 14, the mean being 7.4 and the standard deviation, 2.8.

These figures show that there are over twice as many adult females as adult males. In two of the 28 groups the number of adult males and females was equal; in all other groups the number of adult females exceeded the number of adult males. It should be noted that complemental, solitary males are found but these make up only about 3 per cent of the total population. The factors producing the unequal sex ratio were not determined.

Of interest also is the large number of young animals ob-

served, both infants and juveniles. With almost 50 per cent of the population composed of young animals, one might expect the number of howler monkeys on Barro Colorado Island to be rapidly increasing, and this is substantiated by the fact that the census in 1933 showed a marked gain over that of the preceding year.

3. Group integration, social relations, and intra-group behavior. Within any animal group there are many factors tending to produce both integration and disintegration. The integrating factors include both native propensities and acquired dispositions producing attachments by an individual to a number of other animals and hence to the total clan. A number of the behavior patterns acting predominantly as integrating factors are described below.

The young howling monkey behaves in a parasitic fashion toward the mother, and the mother in turn makes a minimum of positive reactions to the young. Maternal aid and cooperation is, however, evinced in at least three ways: (1) the mother will at times move out of the line of the group progression, go to her young one and posture so that it may easily mount her and be carried, (2) the mother will recover an infant which has fallen to the ground, and (3) she will assist her young in crossing spaces between trees over which the offspring could not cross without aid, either by carrying the infant on her back or by forming a living bridge between the terminals of two opposing branches and allowing the young to walk across her body.

Each female exhibits a specific mode of action toward her own young and each young howler responds to a particular female. Though the bonds between a female and her young gradually weaken, they last even after weaning, as is indicated by the continued association of the juvenile with its mother and a new infant.

Infant howlers develop well-co-ordinated locomotor patterns in the first month of life. Play becomes an important form of activity as soon as they wander from the mother for short periods of time. The amount of play reaches a maximum during the second year of life and then gradually wanes. Play with

other infants appears as an important factor in determining social relations; it includes wrestling and chasing. The play patterns are simple, showing little variation. Play-fighting was observed between late juveniles; but as soon as the severity of play-fighting increases sufficiently to cause pain and frustration, play behavior disappears.

Adult males within the group behave in a peaceful, cooperative manner. Any male that finds a suitable route during arboreal progression will give deep, clucking vocalizations and all other animals will then slowly follow him. Females of a clan usually follow the leading males during group progression, but they may also follow one another. There were no instances observed of fighting or contention between those adult males which were well integrated into a clan and there was no observed competition for sexually receptive females, for food, or for positions. In all defensive situations the males were more active than the females.

There appeared to be no special lasting bonds between any particular male and any particular female. Sexual behavior and choice of partner were determined in large part by proximity rather than by preference. During the period of receptivity, the oestrus period, a single female usually copulated with more than one male. The females gave evidence of being more aggressive than the males in sexual activity. The sexual behavior of the males and females is most accurately described as communal in type, with no indication that it is in any way productive of quarreling or dissension among the males for sexual favors.

4. **Group co-ordination and control.** With observational procedures one is limited in the analysis of those factors operating in group co-ordination and control to a description of the stimulus patterns provided by the movements and the sounds produced by the animals, and the resultant responses of their associates. It is a logical assumption that the various stimulus patterns produced become conditioned stimuli which serve as condensed and meaningful cues to other animals. The degree of foresight on the part of the animal making the original response

can only be conjectured. Four types of control are described: (1) contact control, (2) distant signalization, (3) gesticulation and (4) vocalization.

Contact control is illustrated by the behavior of the mother monkey toward her infant. Originally the mother forces the infant to assume appropriate postures during feeding and during locomotion, but the young howler quickly becomes responsive in these situations to condensed movement cues from the mother.

Distant signalization develops during play. An immature animal may begin running and jumping; others show similar behavior; and through vocal and movement cues their activities are synchronized and co-ordinated. Distant signalization was frequently observed as indicative of danger and as evocative of sexual behavior.

Howler monkeys show a number of stereotyped gesticulatory responses which serve as cues, the most striking example being the rhythmic tongue movements which appear prior to sexual congress.

There can be no doubt that howler monkeys respond specifically to the vocalizations of their associates, probably using between 15 and 20 distinct and important vocal patterns. Typical of these are the following: (1) a deep, hoarse cluck of the lead male to signal progression, (2) rapid, gurgling grunts by males to signal apprehension, (3) a wail, followed by a groan by a female for a fallen young, (4) three little notes or cries by a young who has fallen and (5) a roaring bark to divert the group from feeding, progression, or play in anticipation of the need for defensive actions.

Summary. Howler monkeys live within a definite territorial range, the groups behaving with special reference to definite food and lodge trees. The number of females exceeds the males by over two to one and the total number of adults is little more than the number of infants and juveniles. The intra-group relations are extraordinarily peaceful. Play behavior ceases when play-fighting causes pain. There is no fighting over food, position, or sexual favors. Group co-ordination is effected by

contact cues, gesticulatory signs, and vocal signals. Vocalization is very important, the howler monkeys using up to a score of differential and meaningful vocal patterns.

COMPARATIVE SOCIAL ORGANIZATION OF SUB-HUMAN PRIMATES

In spite of the gulf between man and the lower primates, evolutionally minded anthropologists have theorized concerning the development of the family from lemur to man. In keeping with a Spencerian theory and a North-European folkway, it has been suggested that familial evolution runs the gamut from a promiscuous horde of monkeys to idyllic monogamy among the gorillas.

The truth or falsity of such assumptions must, of course, rest upon the data obtained by controlled observational studies of social behavior made by competent scientists on sub-human primates. Such data as are now available make possible comparison of the howler monkey (an arboreal new-world monkey), the baboon (a terrestrial old-world monkey), and the chimpanzee (an African anthropoid). The anthropoid apes are generally considered to be more advanced than the monkeys, but there are no data to indicate whether new-world or old-world monkeys are the more advanced or higher in the phyletic series. The degree to which the primates chosen are representative of their families or sub-families is an open question.

The observations on the howler monkey in the previously described study were made by Carpenter. The information concerning the baboon is a composite of observations by Zuckerman on the Chacma baboon in the wilds and the Hamadryas baboon in captivity. The chimpanzee was studied by Nissen in French Equatorial Africa.

The baboons are doubtless the most gregarious of all subhuman primates. Although the pack observed by Zuckerman consisted of only 25 animals it appears that baboon groups may number a hundred or even several hundreds. The largest clan of howler monkeys totaled 35 and the largest group of chimpanzees 14, although two instances of temporary combination of groups gave totals of 16 and 18 respectively.

There is every indication that baboon packs are composed of family groups which typically include a male and one or more females, depending upon the male's prowess in fighting and in the maintenance of his dominant status. Thus the social organization of the baboons is that of a group of families with polygyny preferred and monogamy forced upon the less successful males of the society, a system which finds widespread—though often only tacit—acceptance in both primitive and civilized human societies. No family groups exist among the howler clans, which function as the promiscuous hordes for which the anthropologist long sought. But it must be again stated that there is no evidence that this is a more primitive evolutionary form of social organization than that enjoyed by the baboon. The data concerning the chimpanzee are not adequate to permit a definitive answer as to the nature of their social organization. Nissen states that promiscuity is an "extreme possibility," but that it is much more likely that the chimpanzee groups are made up of a polygynous male and his entourage.

In spite of the fact that the baboons are the most gregarious of all sub-human primates, there is evidence of constant strife. The two large groups observed in captivity were decimated by frequent fights for the possession of the limited number of females. Adult baboons, both male and female, killed or captured in the wilds, give evidence through injury or laceration that the family organization is not static nor peaceably maintained. The domestic life of the chimpanzee seems to be a continuous series of quarrels and squabbles. Since, however, most of the evidence is auditory and not visual, Nissen has advanced the "Pollyanna" hypothesis that there may be much noise about trifles even though the noise is not infrequently that of a fear-pain cry. The domestic life of the howler monkey can be criticized only for its non-European pattern and for its saccarine quality, each howling clan being one large and happy family. The moral is obvious, even if distasteful, and suggests that man unfortunately evolved from the wrong primate group.

The data afforded by the observational studies gives no indi-

cation that the mechanisms involved in group co-ordination and control of the howler monkey clans are in any way inferior to those mechanisms operating within the baboon or the chimpanzee social groups.

The limited information extant shows clearly that the patterns of social organization that exist among the various genera of sub-human primates are most varied, and an attempt has been made by Maslow to relate this to basic personality traits. Thus Maslow states that dominance in the baboon is "rough, brutal and aggressive; it is of the nature of a powerful persistent, selfish urge that expresses itself in ferocious bullying, fighting and sexual aggression" and this dominance quality naturally "goes with" the baboon social picture.

Conversely South American monkeys (including the howler) "are generally low in dominance expression and at times it even appears as if dominance is not expressed in their behavior" and this dominance quality "goes with" the social pattern of the howler monkey.

COOPERATIVE BEHAVIOR AMONG PRIMATES

One of the most striking differences that appears from observation of human and sub-human social groups is the large amount of cooperative behavior exhibited by man and the small amount of cooperative behavior shown by the lower animals. Both female howling monkeys and baboons have been observed helping their young during difficult progression, but the cooperation of two adult animals in solving a problem too difficult for one alone is rare indeed. Such facts suggest many intriguing problems. Does the performance of cooperative behavior characterize human social groups and these only? Is failure to exhibit cooperation among sub-primates a result of lack of environmental pressure? Is man the only animal possessing the requisite intellectual capacities for carrying out cooperative behavior? Is cooperative behavior related to the degree or nature of dominance of individuals within the species?

An experimental attack on various phases of the problem of cooperative behavior has been made already by workers at the

Yale Laboratories of Primate Biology. One of these studies is described in the following pages.

The Cooperative Solving of Problems by Young Chimpanzees

M. P. Crawford[4]

The purpose of this study of cooperative behavior among chimpanzees was to investigate and describe the co-ordinated activity of two individuals working for a common incentive. Three different problem situations were used in this investigation. The first was designed to study the stages involved in the acquisition of co-ordinated activity. The second and third problems offered data supplementary to the first and also gave evidence concerning the transfer of co-ordinated activity from one situation to another.

Subjects. Five young chimpanzees, Ross, Bula, Bimba, Kambi, and Alpha, all of whom had previously been tested in various problems in the Yale Laboratories, were used as subjects. All five animals were rated by members of the staff both as to their intelligence and as to their relative dominance status.

Apparatus. The experiments were conducted in a large experimental room with an observation booth built into one corner. Continuous observation of the animals in the experimental room was possible through two one-way vision screens. Near one of these screens was a table upon which electrical control devices were mounted and records were written. A dictaphone was available for recording observations whenever necessary.

The restraining cage for the animals was two meters in length, width and height, and was divided longitudinally into two compartments. A steel grille which allowed the chimpanzees to reach their arms between the bars lined the front of the cage, and a back grille of similar construction formed the rear half of the dividing partition. This latter grille was removable.

The essential apparatus for Problem I consisted of a heavy box which could be drawn toward the experimental cage by one

[4] "The Cooperative Solving of Problems by Young Chimpanzees." *Comp. Psychol. Monog.*, 1937, *14*, No. 2, pp. 1-88.

of two ropes. The weight of the box could be varied so that a horizontal pull ranging from 27 to 180 pounds was necessary to produce movement. Either one or two pieces of food could be placed on the box by means of a hidden, mechanical baiter controlled from the observation booth. The apparatus was so arranged that stylographic records of both the time and the amount of pulling on each rope could be obtained. A similar record of the time that the box first appeared to move was controlled by the experimenter.

The two cord apparatus used in Problem II was so designed that one animal could, by simultaneously pulling down on both cords bimanually, cause a door to fall open and allow food to drop out of a box within his reach.

The handle apparatus used in Problem III was similarly arranged so that one subject alone could push on two handles simultaneously and permit a food tray to swing within his reach.

The apparatus used in Problem I was adapted for the study of cooperative behavior between two animals by the addition of enough weight to make pulling by a single chimpanzee impossible. In Problems II and III the same end was achieved by separating the two cords or handles so far that a single animal could not reach both at the same time.

The procedure used in Problem I gave the subject 15 to 30 minutes of working time. In Problems II and III each session involved 10 trials for training, either individually or in pairs.

In those cases in Problems II and III in which co-ordinated activity began immediately, one of the animals was offered a plaything in order to delay his response and to give his partner an opportunity to use solicitation.

A fundamental difference in training procedure existed between Problem I and Problems II and III in that the individual was trained on only one of the operating mechanisms (one rope) in Problem I but was trained to manipulate both of the operating mechanisms on the other two problems.

Training and development of co-ordinated behavior. Before training for co-ordinated behavior began, all animals

were individually trained on each problem. Following individual solution of the problems, the animals gave no indication of co-ordinated behavior when faced with the problems so changed as to demand co-ordinate behavior for successful solution. This fact is quite in keeping with the observed tendencies of chimpanzees who, when faced with a strange object, employ *alternate, individual manipulation* as the characteristic mode of attack. In an effort to counteract this tendency a series of steps were undertaken to *train* the chimpanzees to utilize a cue that would enable them to time their pulling together. The subjects were first trained in simultaneous activity to pull two light boxes independently at about the same time. This was quickly learned. The heavy box was then introduced, and 5 training steps were used to enable the chimpanzees to respond cooperatively to a common interval cue.

1. The experimenter helped the animals by pulling or pushing the box from front to rear, thus orienting the subjects toward the box and simplifying the task when they pulled.
2. The experimenter accompanied his efforts with the vocal call "pull." He gave this call just before the subjects began their pull on each heave.
3. The experimenter did less and less actual work on the box while, in the apparent response to his call, the subjects did more and more.
4. The experimenter did no pushing or pulling at all, but simply touched the box and called.
5. The experimenter stood away from the box and track and called the animals, who responded simultaneously, moving the box with a series of smooth, regular heaves.

These steps completed the training for stage 1. At the end of this training the animals needed only an external auditory cue to initiate cooperative behavior, and after that they were usually able to pull in their own rhythm and draw the box all the way to the cage in four or five heaves.

Stage 2 in the development of cooperation was characterized by one or both animals watching the other and responding to

the behavior of the partner. The behavior of Bula while associated with Ross may be regarded as typical:

> . . . it was discovered also that she seemed to be responding to . . . the tightening of Ross's rope. . . . (It was usual for Ross to respond first to the verbal cue and to begin work earlier.) This tightening was regularly accompanied by a squeak of the rope. Bula was observed to turn from whatever she might be looking at, take her rope, and begin to pull after one squeak from Ross's rope. The sound probably served as an effective cue. Thus the activity of the partner began to afford a cue to Bula as substitute for the behavior of the experimenter. Gradually Bula became more responsive to the behavior of Ross himself. As training continued, Bula responded to a greater number of the details of Ross's behavior, and to behavior which preceded pulling by longer intervals of time.

The third and final stage in the development of cooperation was spoken of as *solicitation*. Solicitation consisted of attempting to stimulate the partner who did not make any movements preparatory to pulling. This was accomplished by whimpering, hooting or other vocalizations, by beckoning or supplicatory gestures, and by physical contact.

From the above description it is obvious that whereas stage 1 resulted from active tuition and assistance by the experimenter, stages 2 and 3 were learned independently of any outside tuition. Stage 3, particularly, appeared only after other methods of solution had been tried and had failed.

When new pairings were made with a subject or subjects that had already attained cooperative behavior, the preliminary training was curtailed if it appeared that such procedures were likely to prove unnecessary.

Results: Stages of cooperation attained in Problem I. In Table 12 are presented the results obtained with Problem I in each of the 7 pairings of the 5 subjects. From these data it can be seen that all 5 subjects reached stage 1 and in certain combinations, all but Ross showed behavior typical of stage 2. Two subjects, Bimba and Bula, solicited in the manner characterizing stage 3.

TABLE 12

STAGE OF SUCCESS ATTAINED BY 5 CHIMPANZEES IN 7 PAIRINGS TOWARD THE DEVELOPMENT OF COOPERATIVE BEHAVIOR IN PROBLEM I

	BULA		
ROSS	Both reached stage 1. Stages 2 and 3 attained by Bula.		
KAMBI	Both reached stage 1 and 2 and Kambi was solicited by Bula (stage 3).	KAMBI	
BIMBA	Both reached stage 1, 2 and 3.	No co-ordinated behavior observed in 23 sessions, although Bimba solicited Kambi.*	BIMBA
ALPHA	Solicitation by Bula had no effect on Alpha.	Stage 1 and 2 reached by Kambi. Doubtful solicitation by Kambi ineffectual.	Solicitation by Bimba had no effect on Alpha.
	No co-ordinated behavior on Alpha's part.	Alpha reached stage 1 and showed some stage 2 (watching partner) behavior after experimenter's call.	No co-ordinated behavior on Alpha's part.

* Bimba had already solicited Bula, and Kambi had learned to respond to Bula's solicitations.

Results obtained from graphic records of pulling in Problem I. The graphic records of pulling in Problem I gave objective data concerning the amount of pulling by each partner and the time relations of each heave. Each subject tended to show such a characteristic recurrence of patterns of pulling

on practically every trial that the animal might be identified from the record. Thus most of Bimba's stylographic tracings showed a slow rise in tension, maximum tension held only for an instant, and then a rapid fall as if the rope had been released. Bula's records showed a recurrence of a pattern described as a rapid rise in tension with a flattened or rounded top (indicating that maximum tension was held longer than an instant) and then a rapid fall.

The data obtained showed no consistent relationships between amount of pulling and solicitation, or between watching behavior and the relative time of beginning each pull, although it might be expected that the watching chimpanzee would begin to pull later than his partner. There was some evidence in the case of Bimba that pulling alone was replaced by solicitation since after solicitatory behavior developed she was not observed to pull without her partner also pulling. The exact reverse was observed in Bula who was the only other animal to reach stage 3.

Stages of cooperation attained in Problems II and III. The results obtained in the further study of cooperative behavior using Problems II and III are briefly listed in Table 13.

There were only 2 pairings studied in Problem II and 4 in Problem III. Stage 1 was reached in one of the 2 pairings in Problem II and in 3 of the 4 pairings in Problem III. Stage 2 was never reached in Problem II in either pairing and was reached by only 2 of the 5 animals in Problem III. Stage 3, solicitation of the partner to pull, was attained only by Bula and was shown by this animal in but two pairings out of four.

Transfer of solicitational behavior to new pairings. After soliciting behavior developed in the case of both Bula and Bimba in Problem I, these animals solicited all the other chimpanzees with which they were paired in this problem. Thus solicitation first developed in Bula when paired with Ross and was subsequently exhibited in pairings with Bimba, Kambi and Alpha, even though Alpha was unresponsive. Solicitation on the part of Bimba first appeared when paired with Bula and was subsequently transferred to Kambi and Alpha, in spite of

TABLE 13

STAGE OF SUCCESS ATTAINED BY 5 CHIMPANZEES IN 4 PAIRINGS TOWARD THE DEVELOPMENT OF COOPERATIVE BEHAVIOR IN PROBLEMS II AND III

	BULA	ALPHA
KAMBI	Both given individual training on Prob. II, III. Problem III (15 sessions). Co-ordinated and watching behavior developed in both animals but no solicitations could be induced.	Both given individual training on Prob. II, III. Problem II (3 sessions). No co-ordinated nor watching behavior observed. Problem III (22 sessions). Co-ordination usually due to attention to levers, not to partners.
BIMBA	Bula given individual training on both problems, Bimba only on Prob. II. Problem II (15 sessions). Co-ordinated behavior developed but was largely a function of both subjects responding immediately to the problem at the beginning of each trial, rather than from watching the partner. Bula was urged to solicit Bimba but without effect. Problem III (4 sessions). Bula was urged to solicit Bimba (who had no training on this problem) but without effect.	
ROSS	Both given individual training on Prob. II, III. Problem III (9 sessions). Co-ordinated behavior attained by Bula watching and responding appropriately to partner's behavior. No solicitation by Bula.	

the fact that no co-ordinated behavior of either Type I or II appeared between these subjects before solicitational behavior was exhibited by Bimba.

Transfer of solicitational behavior from problem to problem. Cooperation was obtained only after a series of training trials in Problems II and III and its development followed a course similar to that outlined in Problem I.

Solicitational behavior on the part of Bula was transferred from Problem I to Problems II and III but only when paired with Bimba. On the other hand, Bula did not solicit help from Kambi in Problem III under circumstances similar to those in which she did solicit Bimba. Furthermore, Bula would not solicit Kambi on Problem III even on the same day that she was observed to solicit Kambi in Problem I.

Factors influencing solicitational behavior. Five factors other than practice are suggested as influencing the intensity, duration and frequency of solicitation. (1) *Degree of friendship between the partners.* Bimba and Bula were the best friends among the experimental group and showed the most intense, effective and persistent solicitation. (2) *Degree of responsiveness of the partner.* Cessation of solicitation developed when a partner consistently refused to pull when stimulated. (3) *Degree of motivation toward food shown by the solicitor and by the partner.* Loss of food motivation appeared to account for fluctuation in solicitational behavior. (4) *Social dominance relationships.* Neither Bula nor Bimba had any success in soliciting Alpha, the most dominant of the chimpanzees; and Bimba, dominant over Bula, enjoyed greater success in the solicitation of Bula than did Bula in soliciting Bimba. (5) *Intelligence.* Bimba and Bula were rated as more intelligent than either Kambi or Alpha and the first two animals showed far more solicitational behavior than the last two.

Summary and conclusions. Five young chimpanzees were first trained individually in Problem I to pull in a box by means of a rope. The box was then increased in weight so that an individual chimpanzee could no longer move it, but two animals

working together could draw it to the cage and obtain the reward.

In Problem I initial attempts of the paired animals to draw in a box were entirely unco-ordinated. Training in simultaneous activity was next introduced and both members of a pair learned to pull in a light box at approximately the same time. Training was then begun with the single heavy box and the subjects learned to respond cooperatively, guided by visual and verbal cues and by aid from the experimenter. Co-ordinated activity in response to the experimenter's auditory cue characterized stage 1 in learning to cooperate. Stage 2 was reached by the animal when he watched his partner's behavior and timed his pulling with his partner's pull. Stage 3 was attained by an animal when he solicited aid from a partner who was not pulling.

In the first problem all 5 of the subjects attained stage 1 and 3 animals reached stage 2. The 2 animals that showed stage 3 behavior solicited all partners with whom they were paired.

A problem box opened by manipulation of 2 cords was used in Problem II and a box opened by pushing on 2 handles was used in Problem III.

Only 2 of the 5 animals advanced beyond stage 1 in Problems II and III, but 1 of these animals solicited 1 partner (out of 3) in both problems.

These data showed that chimpanzees can respond cooperatively if aided by intensive preliminary training and that solicitation may develop as a method of problem solution. Solicitational behavior among the chimpanzees is probably related to dominance, responsiveness of partner, food motivation, friendship and intelligence.

Part Five

HUMAN DEVELOPMENT

By

BETH L. WELLMAN
State University of Iowa

CHAPTER 15

Inheritance of Mental Abilities

THE INVESTIGATION of the inheritance of mental abilities is a complex matter. While it is known that human beings possess a large number of chromosomes, with possibilities of a wide variety of gene combinations, almost nothing is known about the actual mechanisms of physical heredity in regard to any given human mental ability or complex of mental abilities (such as intelligence). The most obvious reason for this is that it is not possible for the biologist to study human beings under rigorously controlled conditions for such long periods of time as have been found necessary for the determination of hereditary characteristics in lower animals.

When the biologist studies the inheritance of a characteristic in animals he traces its appearance (or non-appearance) through successive generations with animals which are usually carefully selected for known characteristics at the beginning of the experiment. Furthermore, if his study concerns rate of growth, he holds the animals to rigidly controlled environmental conditions; for he recognizes that rate of growth is dependent upon the conditions under which the animals live.

Measurement of intelligence. It can readily be seen that studies of the inheritance of human abilities cannot begin to approximate the necessary rigorousness of experimentation required for clear-cut results. These considerations should perhaps induce us to admit our real ignorance. Because of the complex nature of human abilities, it is unwise to draw inferences about the inheritance of such traits from studies of characteristics which are only roughly similar in the lower forms of animals. Can useful information be obtained, however, from less rigorous evidence? It should be possible to trace the ap-

pearance of a given ability or characteristic through several generations if adequate records have been kept. The practical difficulty here is that the better measuring instruments have been developed too recently. In studying the inheritance of intelligence, it therefore becomes necessary to substitute for the modern intelligence test indirect measures on the earlier generation, such as the socio-economic status of parents.

That the intelligence of children is related to the occupational level and intelligence of their parents has been well established in a number of investigations.[1] Can we then say that this relationship is due to inheritance? Since most of the studies have been made on children who have lived with their own parents, interesting light on this question may be obtained by removing children from their parents to see whether their mental development is affected thereby. The experiments to be described here have done that: they have studied the development of children who would be expected to be low average, dull or retarded in mental ability if they had been brought up by their own parents, but who instead, through force of circumstance, were placed in new homes under better conditions.

Parentage of foster children. A group of 154 infants whose true parents were in the main from the lower socio-economic levels were placed in foster homes which were in general from the average and superior occupational levels. Their mental development was studied by Skodak.[2] Intelligence test scores were available for 80 of the true mothers who were considered representative of the whole group; for the intelligence of their children was similar to that of the children whose mothers' IQ's were unknown. The mean IQ of the mothers was 88; more than half of them (54 per cent) had IQ's below 90

[1] Stoddard, George D., and Wellman, Beth L., "Environment and the IQ," Thirty-ninth Yearbook, National Society for the Study of Education. *Intelligence: Its Nature and Nurture,* Part I. Comparative and Critical Exposition. Bloomington, Ill., Public School Publishing Co., 1940, 405-442.

[2] Skodak, Marie, *Children in Foster Homes: A Study of Mental Development,* Univ. of Iowa Study in Child Welfare, 1939, 16, No. 1, pp. 156.

and 14 per cent were mentally retarded, with IQ's below 70.

Occupational records were available for 110 of the true fathers. Nearly half of them (47 per cent) were unskilled day laborers; while, according to the 1930 census figures,[3] only about 20 per cent of the employed males in the United States are unskilled day laborers. Thus, the fathers of this group of children appeared in the lowest occupational category with more than twice the normal frequency. The percentage of true fathers in the top three occupations was close to that for the general population. On the other hand, the foster fathers' occupations were above the average for the general population: only 1.3 per cent were unskilled day laborers, while the percentage in the top three occupations was far in excess of that for the general population. The accompanying tabulation shows the distribution of occupations for the true fathers and foster fathers in comparison with the 1930 census figures.

TABLE 14

OCCUPATIONAL CLASSIFICATION OF TRUE AND FOSTER FATHERS IN SKODAK'S STUDY

Occupational classification	*General population, employed males, 1930* * %	*True fathers N = 110* %	*Foster fathers N = 154* %
I. Professional	3.1	1.8	13.6
II. Semi-professional and managerial	5.2	4.6	22.7
III. Skilled trades	15.0	14.6	24.0
IV. Farmers	15.3	10.9	25.3
V. Semi-skilled	30.6	10.9	8.4
VI. Slightly skilled	11.3	10.0	4.6
VII. Day laborers	19.5	47.3	1.3

* From Terman and Merrill, *op. cit.*, p. 14.

Intelligence of foster children. The mean age of the children at the time of placement in the foster homes was 2.8 months; no child was older than 6 months. At the time the first

[3] Terman, Lewis M., and Merrill, Maud A., *Measuring Intelligence: A Guide to the Administration of the New Revised Stanford-Binet Tests of Intelligence*. Boston, Houghton Mifflin Co., 1937, pp. xi, 461.

intelligence test was administered the mean age was 2 years. Children under 3½ years were tested on the Kuhlmann-Binet Test, those over 3½ years on the Stanford-Binet. The mean IQ of the children was 116.0, with an S.D. (standard deviation) of 13.6, indicating superior intelligence. Only 3 per cent of the children had an IQ below 90 and none below 80; 69 per cent tested at 110 or above, and 42 per cent tested 120 or above (very superior).

A second test was administered to the entire group after an interval of two years. The mean age of the children was then 4 years, the range being from 2 to 10 years. The mean IQ was 111.5, with an S.D. of 13.2. While their performance was somewhat lower than on the first test, the children were still superior. The question with which we are concerned here is, "Was their performance better than we should expect from children of their true parentage?"

Intelligence of children in relation to fathers' occupation. If we were to compare their mean IQ with that of the child population as a whole at all ages, we would say at once that the foster children were superior in intelligence, since the tests are so standardized that we expect to obtain an IQ of 100 on the average. However, since these children were examined at a very early age, it may be well to ask the additional questions: "Do children at these ages in general test higher than 100?"; "Are IQ's of young children differentiated in accordance with the occupational levels of their fathers?"; and "If so, what mean IQ should we expect from this group?" The answers to these questions can be considered together.

Studies of intelligence of children two to five years of age show that their mean IQ does vary according to the occupational classification of their fathers. For children 2 to 4 years of age tested on the Kuhlmann-Binet Test, Goodenough[4] reported mean IQ's of 116 for the professional group and 112 for

[4] Goodenough, Florence L., *The Kuhlmann-Binet Tests for Children of Preschool Age: A Critical Study and Evaluation.* University of Minnesota, Institute of Child Welfare Monograph Series No. 2. Minneapolis, University of Minnesota Press [c. 1928], pp. 146.

the semi-professional and managerial group. Terman and Merrill give 116 as the mean IQ on the newly revised Stanford-Binet Test for children 2 to 5½ years of age whose fathers are in professional occupations, and 112 as the mean IQ for children whose fathers are in semi-professional and managerial occupations. For 652 children entering the preschool laboratories at Iowa City at ages from two to six years inclusive, with a median at four years, approximately two-thirds of whom come from professional homes, a mean IQ of 117 was obtained by Wellman.[5] Thus, we see that the mean IQ of the foster children in the study described above was, on first examination, almost identical with the mean commonly reported for children from professional homes; and on second examination the mean was very close to that commonly reported for children from the second occupational level.

On the other hand, Goodenough gives 96 and Terman and Merrill give 94 as the mean IQ at the preschool ages for children of day laborers. They give 105 and 104 respectively for the semi-skilled trades. Obviously the IQ of the foster group corresponds more closely to that of the higher occupational group than to that of the lower.

Foster children from the lowest occupational groups. An interesting question which arises in connection with the Skodak material is, "How do children whose true fathers are in the lowest occupational categories compare in IQ with others?" Skodak selected from her total group all children (N = 52) whose true fathers were known to be in the seventh occupational category, that of unskilled day laborer. She found that some had been placed in foster homes representing the upper three occupational categories and some in foster homes representing the lower four categories. Those who were placed in the more privileged homes were higher in IQ at each age than those who were placed in the less privileged homes. Reading from the curve published by Skodak,[6] we find the mean IQ's at ages 2, 3, and 4

[5] Wellman, Beth L., "Iowa Studies on the Effects of Schooling," Thirty-ninth Yearbook, National Society for the Study of Education. Part II. Original Studies and Experiments, 1940, 377-399.

[6] *Op. cit.*, p. 83.

years were approximately 118, 116, and 113, respectively, for the children who were placed in the superior foster homes. These children were thus approximately 20 points higher than children from the same true father occupational level who live in their own homes (if the Goodenough and Terman and Merrill figures are representative).

The other children whose true fathers were in the seventh occupational category but who were placed in foster homes representing the lower four occupational groups had mean IQ's of approximately 110, 107, and 104 at ages 2, 3, and 4 years. They were thus approximately 10 to 15 points higher in IQ than children from the same true father occupational level who live in their own homes.

Children of feeble-minded mothers. Another comparison, perhaps even more interesting, in the Skodak material, was that dealing with children whose true mothers were known, by test and by social criteria, to be feeble-minded. There were 16 such children in the group. The mothers were high-grade mental defectives, having a mean IQ of 66. After placement in foster homes, the mean IQ of the children on their first test, at an average age of 2 years, was 116, almost identical with the mean for the total group. Again, the children who were placed in foster homes representing the upper three occupational categories were higher in IQ than those placed in foster homes representing the four lower occupational categories although the IQ's of the true mothers were slightly higher for the latter group. These data are summarized in Table 15.

TABLE 15

	8 children placed in homes representing three upper occupational categories	*8 children placed in homes representing four lower occupational categories*
Mean IQ on first test	118.6	113.6
Mean age on first test	41 months	20 months
Mean IQ on second test	112.3	104.0
Mean age on second test	69 months	51 months
Mean IQ of true mothers	63.9	68.9

In interpreting these figures the question arises, "What level of IQ in children should we generally expect when the mothers are feeble-minded?" Though the literature does not provide a basis for a good estimate, we should not ordinarily expect such children to be superior in intelligence, as these foster children were.

Some information about an additional group of children who were taken away from their feeble-minded mothers has recently been published by Speer.[7] He studied 68 physically normal children whose mothers were feeble-minded. All but one mother was institutionalized. The mean IQ of the mothers was 49.0, with a range of 38 to 64. The children were removed from their mothers when the home was broken up and placed in boarding homes (not in adoptive foster homes as the children studied by Skodak had been). Speer found that there was an inverse relationship between the children's IQ and their age at removal from the home, their age representing the number of years spent in their own inferior homes. The younger children were normal in intelligence, while the older children (12 to 15 years) were feeble-minded to about the same degree as the mothers. Not one of the children placed before the third birthday had an IQ below 90 and no child placed after the twelfth birthday had an IQ above 70. The median IQ in relation to age of placement (or number of years in their own homes) is given in the accompanying tabulation.

TABLE 16

Age at placement, years	*Number*	*Median IQ*
0 to 2	12	100.5
3 to 5	19	83.7
6 to 8	12	74.6
9 to 11	9	71.5
12 to 15	16	53.1

[7] Speer, George S., "The Mental Development of Children of Feeble-minded and Normal Mothers," Thirty-ninth Yearbook, National Society for the Study of Education. *Intelligence: Its Nature and Nurture*. Part II. Original Studies and Experiments. Bloomington, Ill., Public School Publishing Co., 1940, 309-314.

In the general population such decreases in IQ with age are not expected. However, we might raise the question, "Do children of feeble-minded mothers generally decrease in IQ with increasing age?" From the standpoint of implications for heredity, we might also ask, "If so, is this decrease inevitable because constitutional?" A partial answer may be found by repeatedly testing the children of feeble-minded mothers who have been removed from their own homes. Skodak found no indication of a greater decrease in IQ on successive tests of the children of feeble-minded mothers up to age six than for the foster children whose mothers were not feeble-minded.

Finally, we might ask, "How did the parents of these children come to be feeble-minded?" "Have we any evidence that their feeble-mindedness was genetic in origin?" It is conceivable that some, at least, of these parents might have been higher in intelligence if they had been brought up under different circumstances.

Summary and conclusions. It has been pointed out that there have been no carefully controlled experiments based on selective breeding of human family lines designed to investigate the possible inheritance of mental abilities; and that we should be extremely cautious in reasoning by analogy from experiments on the inheritance of structural characteristics in the lower animals.

As a usual thing, children whose fathers are in the lowest occupational categories have lower IQ's than children whose fathers are in the higher occupations. It was shown, however, that when children were removed from underprivileged homes at an early age and placed in homes representing superior occupational levels, their mental development was correspondingly affected. A group of children whose parents were of inferior mental and social status, but who had not lived with their own parents, was found to be superior in intelligence.

It appears that children of feeble-minded parents, if brought up by those parents, will tend to be mentally retarded in later childhood, while the prognosis is good for average or superior intelligence if the children are removed from their parents at an early age and brought up under favorable circumstances.

Environmental Stimulation of Mental Ability

CONTROVERSY concerning the nature and changeability of the IQ has reached an acute stage. Basically, the two main issues in the controversy are : (1) whether the IQ changes in substantial amounts or is relatively constant; and (2) the extent to which changes, when found, can be attributed to environmental conditions.

Cause of IQ change. To persons who believe that the IQ is a direct representation of native endowment alone, the idea that it changes is usually disturbing, for the proof of native endowment on which they lean most heavily is a belief that the IQ remains constant. Therefore, when changes in IQ are brought to their attention, such persons look for explanations in the inadequacies of the test. For children examined at the age of six years and above, the explanation of large individual changes most often proffered by them is that the administration of the test was inadequate, either because of some unusual condition of the child at the time of one of the examinations, or because of the inexperience of the examiner. For children examined at the preschool ages, an additional question is sometimes raised as to adequacy of the standardization of the test.

The experiments reported below should leave little doubt that the IQ does change, both for individuals and for groups. Two illustrations are given of large group increases in IQ and two of large group decreases.

INCREASES IN IQ AT PRESCHOOL AGES AND DURING LATER CHILDHOOD

Relation to preschool attendance. Children attending the preschool laboratories of the Iowa Child Welfare Research Station increase in mean IQ from fall to spring, but fail to gain over the summer vacation months. As a group, these children are mentally superior and they come from homes of high socio-economic status, approximately two-thirds of their fathers being engaged in professional occupations. Although the mean IQ for 652 children on entrance to preschool was 117, which places them in the "superior" category,[1] in an average interval of six months of preschool attendance they gained 6.6 IQ points. On the other hand, no real change was shown over the six months which included the summer (actually a loss of 1.5 IQ points in 432 cases).

Over a period of two years of preschool attendance, changes in IQ have been studied for 228 children whose mean initial IQ was 117 at a mean age of 40 months. The changes in IQ for these 228 children were as follows:

	Change in IQ
First year in preschool	+ 7.0
Summer	− 0.4
Second year in preschool	+ 3.8
Net change	+10.4

These children gained in IQ over their first year in preschool, did not change appreciably over the summer, but gained again during their second year in preschool. At a mean age of 60 months they were 10 points higher in IQ than on their initial test a year and a half earlier.

Sixty-seven children have been studied over three years of preschool attendance. The number of children was necessarily much smaller in this group than in the other groups, owing to the fact that all children entering at four and five years of age

[1] Wellman, Beth L., "Iowa Studies on the Effects of Schooling," Thirty-ninth Yearbook, National Society for the Study of Education. *Intelligence: Its Nature and Nurture.* Part II. Original Studies and Experiments. Bloomington, Ill., Public School Publishing Co., 1940, 377-399.

had to be excluded because they were too old to attend preschool three years. If a child was absent from preschool because of illness or for any other reason at any one of the six testing periods in the three years, he was excluded from the comparison. Losses were sustained from this group, too, when parents moved out of the city or did not continue to enroll the child. Although there may possibly have been some unknown selective factors in this group, comparison with the other groups shows that the changes over the first two years and over the summer were very similar to those made by the larger groups.

The mean IQ of the 67 children on entrance to preschool was 117 at a mean age of 32 months. They were 64 months old on the final test. All of their mean gains were made during periods of preschool attendance, while in each summer they showed a slight negative change.

	Change in IQ
First year in preschool	+ 7.7
Summer	— 0.6
Second year in preschool	+ 4.3
Second summer	— 2.6
Third year in preschool	+ 1.7
Net change	+10.5

Increases during elementary school. Are changes peculiar to the preschool ages, or do significant changes occur also at the elementary school ages? Changes over a period of 4½ years were studied for a group of 58 elementary school children, all of whom were continuously enrolled in the University of Iowa Elementary School and none of whom had attended preschool. The test used was an individual one, the Stanford-Binet. This was first administered shortly after the children's entrance to first grade at a mean age of 70 months, at which time the group had a mean IQ of 115, or superior intelligence. Retested at the age of 10 years (124 months), they showed a mean IQ of 126, having made a gain of 11.6 IQ points.

There is evidence to show that gains such as these cannot be explained by the number of tests taken, that is, by becoming "test wise." For example, in another study differences in the

amount of IQ change were found for two groups of school children initially equal in IQ who received the same number of tests but attended different schools, and the brighter children did not show the greatest increases.

DECREASES IN IQ AT PRESCHOOL AGES AND DURING LATER CHILDHOOD

Controlled preschool experiments. In an orphanage setting, changes in the intelligence of children at preschool ages were studied.[2] All children in residence in the orphanage who were of preschool ages at the time the project began were divided into two matched groups, one of which was to attend preschool while the other continued with the usual orphanage routine. The groups were matched for IQ, age, length of orphanage residence, sex and nutritional status. The children who were committed to the orphanage later were assigned to the two groups in such a manner as to keep them as nearly equated as possible. One group attended preschool in a special building on the orphanage grounds for approximately five hours a day, five days a week. When not in preschool, the children were in the same environment as the non-preschool, or control group. The control children spent most of their time in a "cottage" with approximately thirty to forty other children of the same ages, under the supervision of an untrained matron, assisted by one or two adolescent orphanage girls. They had access to practically no play materials or equipment.

There were 21 preschool children and 21 control children who remained in residence for at least one year after the experiment started. Both groups were classified as dull-normal initially. The 21 control children showed a mean loss of 6.1 IQ points in slightly less than two years' time,[3] their mean IQ

[2] Skeels, Harold M., Updegraff, Ruth, Wellman, Beth L., and Williams, Harold M., *A Study of Environmental Stimulation: An orphanage Preschool Project.* University of Iowa Studies in Child Welfare, 1938, 15, No. 4, pp. 191.

[3] This particular analysis was not reported in the monograph cited above, but may be found in an article by Beth L. Wellman and Edna Lee Pegram, "A Study of Environmental Stimulation in an Orphanage

dropping from 80.7 to 74.6 while their mean chronological age increased from 32.6 months to 55.6 months.

Since some of the preschool children showed poor attendance, and since, when not in preschool, they were in the control group unless ill enough to be in the hospital, it was important to take into consideration the proportion of days that they attended preschool. The 21 preschool children were therefore subdivided into two groups: (1) those who were in preschool at least as many days as not, and (2) those preschool children who actually were in the control group more days than they attended preschool. There were 13 preschool children with an attendance of 50 per cent or better. They gained 6.8 IQ points, their mean IQ increasing from 82.6 to 89.4, while their mean chronological age increased from 35.8 to 57.8 months. The other 8 preschool children, who were in preschool fewer days than they were in the control group, lost 3.7 IQ points, while their mean chronological age increased from 31.5 to 48.4 months. Their change in IQ was between that exhibited by the control group and that of the other preschool group.

The difference in IQ between the control children and the 13 preschool children with the better attendance records was 1.9 points initially, but 14.8 points finally. The final mean for the control children fell in the borderline classification between feeble-mindedness and dull-normal, while the final mean for these more regular preschool children was at the high end of the dull-normal classification. Initially four of the control children and three of the preschool children tested 70 or below in IQ. On the final test, nine of the control children and no preschool child tested 70 or below in IQ.

TABLE 17

Orphanage groups	*Number*	*Initial IQ*	*Change in IQ*	*Final IQ*
Control	21	80.7	—6.1	74.6
Preschool with less than 50% attendance	8	83.5	—3.7	79.8
Preschool with 50% or more attendance	13	82.6	+6.8	89.4

Preschool Project Re-evaluated in Terms of IQ Changes Made by Individuals," to be submitted for publication shortly.

Progressive decreases among mountaineer children. Large decreases in IQ in later childhood have been reported in five studies in isolated mountain communities However, in these studies the same children were not tested at two different age levels in any one community, but different children were tested at the different age levels. The communities sampled were located in Kentucky, Georgia, Virginia, and Tennessee. A number of different tests were used, usually of the group type. The studies [4] are unanimous in reporting large and progressive decreases in IQ with advancing chronological age. In every study the mean or median IQ in early childhood (5 to 8 years) was either average or dull-normal, and the mean or median in later childhood represented feeble-mindedness or borderline intelligence. The decreases in means (or medians) ranged from 12 to 38 IQ points, as is shown in Table 18.

TABLE 18

Investigator	*Chronological age—years*	*IQ*	*Chronological age—years*	*IQ*	*Difference in IQ*
Hirsch	5 and 6	87	14	75	—12
Asher	7	84	16 and older	65	—19
Wheeler	6	95	16	74	—21
Sherman & Key	6 to 8	84	14 to 16	52	—32
Edwards & Jones	7	108	15 and above	70	—38

Summary and conclusions. The experiments described have shown systematically large increases or decreases in IQ both at the preschool and later childhood ages. Inspection of the results will reveal that explanations for such changes in

[4] Asher, E. J., "The Inadequacy of Current Intelligence Tests for Testing Kentucky Mountain Children," *Ped. Sem. & J. Genet. Psychol.*, 1935, *46*, 480-486.

Edwards, A. S., and Jones, Leslie, "An Experimental and Field Study of North Georgia Mountaineers," *J. Soc. Psychol.*, 1938, *9*, 317-333.

Hirsch, Nathaniel D. M., "An Experimental Study of East Kentucky Mountaineers: A Study in Heredity and Environment," *Genet. Psychol. Monog.*, 1928, *3*, 139-244.

Sherman, Mandel, and Key, Cora B., "The Intelligence of Isolated Mountain Children," *Child Develop.*, 1932, *3*, 279-290.

Wheeler, L. R., "The Intelligence of East Tennessee Mountain Children," *J. Educ. Psychol.*, 1932, *23*, 351-370.

terms of test inadequacies are not satisfactory. Inexpertness of the examiners and unusual condition of the children at the time of examination are ruled out as explanatory factors because of the large number of children tested, the number of examiners in the various studies and the care taken in research investigations to obtain adequate tests. Inadequacies in standardization of the tests cannot explain increases in one group and decreases in another group at the same age levels.

It will be noted that in the experiments described various types of controls have been used. In the studies of Iowa City preschool children, self-groups were compared; that is, the children were compared with themselves for changes in IQ during different periods known to differ in environmental impacts. In the orphanage preschool study, groups which were similar in initial ability and which experienced the same general environment but differed with respect to preschool attendance were compared. Though specific control groups were not used in the studies of the non-preschool children attending the University Elementary School or in the studies of mountaineers, the results obtained for them can be compared with the ordinary expectancies at these ages. In a sense these groups serve somewhat as controls to each other, since the initial tests fell in the same age ranges.

It seems to the writer that the only explanation that logically applies to all of these results is that the IQ changes when changes are made in the environment, and that when the environment differs markedly, over an extended period of time, from that experienced by children in general, the IQ is correspondingly affected. There are environments under which the IQ remains relatively constant, of course. Indeed it is probable that more children experience such environments than experience the markedly varying environments dealt with in the studies just described.

CHAPTER 17

Personality and Social Behavior

WHY DO children differ in personality and social behavior? Do children show consistent and measurable behavior patterns? Can these patterns be changed, and by what means? What are the influences that shape child personality and social behavior?

With the hope of throwing some light on these questions, two investigations in which the behavior of young children was experimentally changed will be described here. Both experiments used training programs aimed at changing the behavior of the children who were initially low on test measures (see below). The first was successful in modifying the children's behavior towards other children and the second was successful in modifying the children's behavior when they were faced with a difficult task. A third experiment described here deals with the effect upon the social behavior of ten-year-old boys of the attitudes expressed by the adults who supervised their activities.

Ascendant behavior of preschool children. A tool for measuring ascendant behavior of preschool children was developed by Jack,[1] who included two types of behavior in the concept of ascendance: (1) pursuing one's own purposes against interference, and (2) directing the behavior of companions. The children were taken in pairs to a room containing a sand box with attractive toys, where they were left alone while their behavior was observed through a one-way vision screen. Each child was paired for a five-minute period with each of ten other children from his own preschool group. The items of ascendant

[1] Jack, Lois M., "An Experimental Study of Ascendant Behavior in Preschool Children," in Jack, Lois M., Manwell, Elizabeth Moore, Mengert, Ida Gaarder, and others, *Behavior of the Preschool Child*, Univ. Iowa Studies in Child Welfare, 1934, 9, No. 3, pp. 171 (pp. 7-65).

behavior observed were: attempts to secure materials (verbal or forceful); success in securing materials; defense of materials; attempts to direct companion's behavior; companion's compliance to directions; the forbidding, criticizing or reproving of companion; and the providing of a pattern which the companion imitated. A child's ascendance score was the total number of instances of such behavior in the ten pairings.

Self-confidence and ascendant behavior. Jack secured scores on two groups of four-year-old children. On the basis of these scores the children were divided into thirds, those receiving the lowest scores being considered nonascendant. The five nonascendant children in one group of 15 children were then selected for a special training program. After careful observation of the children in the preschool play group, it seemed that the most obvious characteristic of the nonascendant children was lack of self-confidence. Consequently, the training program aimed to increase their self-confidence by giving them mastery over certain materials and situations. Three training materials were used, a story, a mosaic puzzle, and a picture puzzle. The experimenter worked with each child individually until he became thoroughly familiar with the materials and had mastered the skills involved.

Training of nonascendant children. After the child had mastered the training materials, other members of the preschool group were brought in, to whom the materials were entirely new. In this arrangement the trained child thus had the advantage over the other children of previous acquaintance with the situation and the materials. Following these pairings, ascendance scores were again obtained by the same method as in the initial measures, using the original sand box and toys.

The five nonascendant children who had been trained showed a large increase in mean score from their initial measure to the final ascendance measure. Their final mean score was nearly equal to the mean score for the remainder of the group.

	Number	*Initial ascendance score*	*Final ascendance score*	*Difference*
Trained	5	38.4	77.8	39.4
Untrained	10	79.8	87.0	7.2

Using a similar approach, Page [2] trained two groups, one of nonascendant three-year-olds, and the other a group of three- and four-year-old children composed of six children, two of whom were initially nonascendant, two moderately ascendant, and two ascendant. Since Page found that adequate scores could be obtained from five pairings, the ascendance scores throughout are lower than those obtained by Jack from ten pairings.

Page, using seven nonascendant three-year-olds as subjects, was successful in increasing the scores in every case. The eleven children constituting the remainder or control members of the preschool group showed a slight loss.

	Number	*Initial ascendance score*	*Final ascendance score*	*Difference*
Trained	7	9.4	28.4	19.0
Untrained	11	27.4	23.6	—3.8

In the second group trained by Page, the two nonascendant children and the two moderately ascendant children increased in score. One of the trained ascendant children changed only one point and the other lost 13 points; in the case of these two children mastery of the training materials was not reflected in an increased ascendance in the sand box situation. The seventeen children constituting the remainder of the preschool group showed a very slight change, a loss of —1.3 points.

	Number	*Initial ascendance score*	*Final ascendance score*	*Difference*
Trained nonascendant	2	12.5	25.5	13.0
moderately ascendant	2	27.0	43.0	16.0
ascendant	2	37.0	31.0	—6.0
Untrained	17	22.8	21.5	—1.3

Reaction to failure. The behavior of preschool children when confronted with a difficult task was observed and meas-

[2] Page, Marjorie Lou, *The Modification of Ascendant Behavior in Preschool Children,* Univ. of Iowa Studies in Child Welfare, 1936, 12, No. 3, pp. 69.

ured by Keister.[3] Two test situations were used, a puzzle box and a weighted box. In the puzzle box were several objects of interest to children, such as a sailboat and an engine, which were fitted closely into the available space. The experimenter removed the objects and invited the child to put them back into the box so that the lid could be closed, allowing fifteen minutes for this performance.

The behavior of the child was timed and recorded item by item. From a group of eighty-one children three to six years of age, 15 children whose behavior was judged to be immature were selected for special training. A child was judged to be immature if he showed two or more of the following reactions: (1) giving up attempts to solve the puzzle box in less than five minutes, (2) requesting help during more than one-half of the total time of the test, (3) manifesting destructive behavior, (4) making more than two rationalizations, (5) evidencing exaggerated emotional responses.

Training "immature" children. The training was aimed at teaching the child to persist longer, to depend less on the adult for help and to see a problem through with composure. For this a series of problems consisting of four picture puzzle books, progressively more difficult, was presented. The experimenter read the story to the child and, as she reached the part illustrated, stopped for the child to put the puzzle together. The second training situation consisted of building a block boy that stood three feet high. The construction was difficult to make steady, usually necessitating several attempts in order to complete it.

Twelve of the 15 "immature" children completed the training. When, following the training, they were retested on a new form of the puzzle box test, their behavior showed statistically significant improvement in number of attempts to solve alone,

[3] Keister, Mary Elizabeth, "The Behavior of Young Children in Failure: An Experimental Attempt to Discover and to Modify Undesirable Responses of Preschool Children to Failure," in Updegraff, Ruth, Keister, Mary Elizabeth, Heiliger, Louise, and others, *Studies in Preschool Education I*, Univ. of Iowa Studies in Child Welfare, 1938, 14, pp. 282 (pp. 27-82).

interest and control of crying. Such emotional behaviors as sulking, crying, destructive behavior, yelling, and motor manifestations of anger dropped out entirely; there were also decreases in whining, asking another to solve, and rationalizations. The mean number of minutes in which a given type of behavior was manifested before training by the 15 "immature" children and after training by the 12 "immature" children who completed the training is shown in Table 19. The pattern of behavior after training was closely similar to that of the remainder of the group.

TABLE 19

MEAN OF RESPONSE IN MINUTES ON PUZZLE BOX TEST

Behavior	*Before training* $N = 15$	*After training* $N = 12$	*Remainder of group* $N = 54$
Attempts to solve alone	8.5	11.2	13.0
Interest	6.0	11.0	12.4
No overt attempt	6.0	2.1	1.6
Asks another to solve	3.6	.5	.8
Rationalizes	2.8	1.6	1.0
Whines	2.6	1.0	.5
Asks help	2.5	2.4	1.5
No emotional manifestations	2.5	1.3	1.7
Cries	1.7		
Indifference	1.0		.1
Sulks	.8		
Destructive behavior	.6		
Yells	.3		
Motor manifestations of anger	.2		
Smiles	.03		.3
Laughs		.1	.1

Influence of attitudes of club leaders. The behavior of ten-year-old boys towards each other in a club situation has been shown to be affected by the attitude adopted by the adult leader supervising their activities. Four clubs of boys, meeting once a week for a six weeks' period, were studied for three periods each under different kinds of leadership. Five groups had "democratic" leadership, five "autocratic" leadership, and two "laissez-faire" leadership; four adult leaders participated. Each group came under the influence of both democratic and

autocratic leadership, and each leader adopted the autocratic and the democratic rôles at least once.[4]

In the "autocratic" groups, all determination of policy was made by the leader; techniques and activity steps were dictated one at a time, so that future steps were uncertain to a large degree; the particular work task and work companion of each member was dictated by the leader; and the leader was personal in his praise and criticism of the work of each member and remained aloof from active group participation.

In the "democratic" group all policies were determined through group discussion and decision, with the assistance of the leader. Activity perspective was gained during the first discussion period and general steps toward the group goal were sketched. Where technical advice was needed the leader suggested several alternative procedures. The leader was objective and fact-minded in his praise and criticism.

In the "laissez-faire" group the leader did not participate in the determination of policy. He supplied materials and made it clear he would supply information when asked but took no part in work discussions and made no attempt to participate in activities or interfere with the course of events.

Aggressive and apathetic behavior. The amount of aggressive behavior was found to vary with the different conditions of adult leadership. For the purpose of this experiment, aggression was defined as any reaction, verbal or physical, that was "hostile" or "joking hostile." In the democratic groups there was a moderate amount of aggressiveness, in the laissez-faire groups it ran high, and in four of the autocratic groups it was very low or apathetic, but in one autocratic group it was high. The average number of aggressive actions per meeting was as follows:

Democratic	20
Laissez-faire	38
Autocratic "apathetic"	2
Autocratic "aggressive"	30

[4] Lewin, Kurt, Lippitt, Ronald, and White, Ralph K., "Patterns of Aggressive Behavior in Experimentally Created 'Social Climates,'" *J. Soc. Psychol.*, 1939, *10*, 271-299.

The same autocratic adult leader who obtained the high aggressive responses in one group was an autocratic leader of another group in which apathetic behavior was obtained. Apathetic behavior was described as dull, lifeless, repressed behavior, with little smiling, joking or freedom of movement.

The accompanying tabulation shows the type of leadership provided in the three periods for each club and the amount of aggressive behavior which obtained. Clubs 1 and 2 had different leaders in each period; clubs 3 and 4 had the same leader in their first and third periods.

Club	*First period*	*Second period*	*Third period*
1	Democratic-moderate	Autocratic-low	Democratic-moderate
2	Autocratic-low	Democratic-moderate	Autocratic-low
3	Laissez-faire-high	Autocratic-high	Democratic-moderately high
4	Autocratic-low	Laissez-faire-high	Democratic-moderate

The children registering apathetic behavior under autocracy showed outbursts of aggression on days of transition to a freer atmosphere; there was also a sharp rise in aggression when the leader left the room. On being interviewed, 19 of the 20 boys said they liked the democratic leaders better than the autocratic ones, and seven out of 10 liked the laissez-faire leaders better than the autocratic leaders.

Two "wars" were engaged in between two club groups that met in the same large room. Both "wars" occurred after the departure of a hostile stranger, the stranger being a graduate student who played the rôle of janitor or electrician and who criticized freely the group's work efforts.

Summary and conclusions. Experimental attempts to modify certain personality and social behavior patterns have been described. The studies were concerned with increasing the ascendance scores of preschool children, improving "immature" reactions to failure among preschool children and changing the amount of aggressive behavior among ten-year-old boys.

Three major conclusions arise from these experiments:

1. Certain characteristics of the young child's behavior in relation to other children can be changed by training designed to change his evaluation of his own status.
2. Similarly, the young child's personality as revealed by work habits can be changed by suitable training.
3. The attitude of children toward others in a group is colored by the way the supervising adult treats the group.

Physical Maturation and Motor Achievements

MATURATION is a term applied to the progressive development of the organism as the child grows older. There are certain stages or landmarks of physical and motor development through which all children seem to pass in the same sequence; for example, infants sit up without support before they develop locomotion, and some form of locomotion (creeping, crawling, hitching, etc.) precedes upright walking. Because of the uniformity with which these and similar phenomena usually occur, some persons have been led to think of maturation as having a fixed, predetermined and invariable course, and some have included in their concept of maturation the idea that the rate or speed of development is also predetermined. Such persons hold that manipulation of experience or environment, within the ordinary ranges, produces little effect upon the process of development. Others believe that rate of development may be altered considerably by environmental conditions.

Individual differences. That there are wide individual differences in rate of acquisition of physical and motor skills has been well established. For example, the age at which children walk alone varies greatly: the median age reported in the literature is approximately 15 months, but in a study of a group of one thousand children of extremely high mental ability,[1] the ages at which the children were reported to walk alone ranged from seven months to over 24 months.

[1] Terman, Lewis M., *Genetic Studies of Genius*. Vol. I. *Mental and Physical Traits of a Thousand Gifted Children*. Stanford University, California, Stanford University Press, 1925, pp. xv, 648.

Practice versus training. The extreme maturationist holds that individual differences have been or are very little affected by external conditions or training. It is true that studies of the effects of repetitive practice upon the acquisition of motor skills of young children have revealed little effect from practice *of the amount and kind utilized in the investigations.* But we must be careful not to confuse practice with training, and further we must be careful not to generalize from one method of training or set of experiences to other methods or experiences.

Practice may be defined as doing something over and over again. Rather uniformly studies have shown that simple repetitive practice has little effect upon the acquisition of motor skills by young children. For example, an experimental group of 30 preschool children given practice for several weeks in throwing a ball at a moving target [2] did not improve in marksmanship noticeably more than a control group of 30 preschool children who did not experience this systematic practice.

Real training, on the other hand, includes specific instructions and suggestions of technique progressively adapted to the stage of mastery of the skill. The most effective training is therefore individualized, taking into account the readiness of each individual for the next step in the process and the best means by which he may reach the desired goal with maximum efficiency and speed. A general example of training is the coaching of teams for competitive games; here differences in training methods enable one football coach to turn out a winning team while another, working with similar material, consistently produces teams which lose.

Measurement of capacity. The person who believes that training is effective will be cautious in making statements about the capacity or potentiality of any given individual, because he realizes that the measurement of ability at any given time

[2] Hicks, James Allan, *The Acquisition of Motor Skill in Young Children: An Experimental Study of the Effects of Practice in Throwing at a Moving Target,* Univ. Iowa Studies in Child Welfare, 1931, 4, No. 5, pp. 80.

may reflect in part the type of training the individual has received. Perhaps the individual under consideration would have attained a higher level of achievement if another method of training had been used. We thus cannot be entirely certain exactly when an individual has reached his top limit or capacity. We can say only that, under the conditions that have prevailed, he has shown the degree of achievement obtained.

There are, of course, limits of achievement beyond which no one is likely to go. We could therefore estimate an upper limit that would apply to all persons, but that would not be helpful in determining how far short of this limit the capacities of any given individual fall. We might conclude that we have measured capacity when we have determined the achievement attained by an individual after maximum or optimum conditions of training, but this would assume that full knowledge of what constitutes optimum training for the individual in question is available. In view of these difficulties, it seems wiser to speak in terms of *achievement* rather than in terms of capacity.

Specificity of motor abilities. Motor abilities are generally found to be rather specific; that is, individuals who are high in one motor skill may or may not be high in another. Since this is the case, it is not to be expected that training which has changed an individual's status in one skill will necessarily carry over or transfer to other skills.

Experimental studies in training. The following experiment demonstrates some effects of an unusual and long-continued training program upon the rate of acquisition of motor skills by a child. The exercise program [8] for this child was begun when he was 20 days old. He spent approximately 35 hours a week in the laboratory, the remainder of his time being spent at home with his father and mother, a twin brother, and five other brothers and sisters. During the early months he was exercised for a short period at two-hour intervals; later the exercise was more extensive, occupying about three hours a day.

[8] McGraw, Myrtle B., *Growth: A Study of Johnny and Jimmy,* New York, D. Appleton-Century Co., 1935, pp. xxi, 319.

Although McGraw compared the development of the experimental child with that of his twin brother who did not undergo the special training, these comparisons will be omitted here for the sake of brevity. Rather, the comparisons will be between the experimental twin and children in general, in an attempt to answer the question "How did the experimental twin compare in ability with what is ordinarily to be expected of children of these ages?"

Whenever the effects of training upon development are evaluated, it is necessary to make an estimate or prediction of what the development would have been without the training. This can be done more accurately for group averages than for single cases. Since individual differences are so great, and little is known about the causes of these differences, prediction of the development of motor achievements of an individual cannot now be made with accuracy. In other words, it is difficult to tell where within the normal distribution the performance of a given individual is likely to fall. This state of affairs renders uncertain the interpretation of differences between two individuals whose achievements are within the ordinary ranges. However, if a child has been subjected to unusual experiences and his achievements are extraordinary, this combination of events constitutes reasonable justification for formulating the hypothesis that there was a causal connection between the experiences and the achievements. The probability of obtaining such a combination by chance is extremely small. It seems to the writer, therefore, that the most crucial comparison in McGraw's study is between the experimental twin and ordinary expectancy for children of this age.

Out of the array of achievements studied the following have been selected for comment here: creeping up and down inclines, walking up and down inclines, getting on and off stools and roller skating. The main technique in training was watching for the emergence of an ability and capitalizing on it by encouraging its exercise. For example, creeping was encouraged by placing lures ahead of the child out of his reach.

When the child was in the creeping stage, the experimenter

endeavored to elicit more difficult creeping by inducing him to ascend and descend inclines of varying steepness. For this purpose, a 72-inch slide, 16 inches wide, was placed at the following degrees of incline: 11°, 18°, 24°, 32°, 40°, 48°, 61° and 70°. Daily practice was begun on the 11° incline when the child was 8 months old, at which time he was just beginning to develop associated creeping movements. When he was 12 months old he crept up the 61° incline without assistance and when 22 months old he ascended the 70° incline with what is described as "incredible facility." The experimenter states that no other child under 24 months observed by her succeeded with inclines of 61° or 70°.

Daily practice in walking up and down inclines was begun at 11 months, when the child had been walking only a short while. At 12 months of age he walked up an incline of 18°, at 14 months an incline of 24°, at 24 months an incline of 40°.

When the child was 9 months old, daily exercise was begun in getting on and off stools varying in height from 7½ inches to 63¼ inches. The child was given help and verbal instructions on the method of turning to get off the stools. At 13 months he had developed an adequate method of getting off stools higher than his body and at 15 months he began getting off the 63¼ inch stool.

Perhaps the most spectacular of his skills was that of roller skating. When 12 months of age he had his first experience on non-ball-bearing roller skates. Within a few days the notion of pushing his feet rather than stepping had been established, and at this time he was placed on four-wheel ball-bearing skates. At 15 months he began to move the right and left foot alternately and was capable not only of steering a straight course but of steering around curves even when coasting at a rapid rate. When 17 months old he was able to coast 188 feet on a gradual slope of 3.1°, which he soon learned to coast down backward. At 20 months he was placed on two-wheel skates, and in less than a month he was able to travel 8 or 10 feet across a cement floor. At 24 months when coasting backward

he could stoop over, place his hands on his feet and see to steer by looking between his legs.

Summary and conclusions. The experiment described has demonstrated that considerable acceleration in the acquisition of certain motor skills may be accomplished by appropriate exercise and training.

We have not been concerned here with the permanence of retention of these skills; that is another matter. In the published account [4] of the development of the twins at the age of six years the author states that the degree of retention of a motor performance is partly contingent upon the state of fixity the activity had attained at the time practice was discontinued. In addition she states that the following factors appear to play an important role in determining the permanency or the deterioration of a motor skill: (1) amount of continued practice (or lack of it); (2) the child's attitude towards the performance; and (3) changes in body proportions.

[4] McGraw, Myrtle B., "Later Development of Children Specially Trained During Infancy: Johnny and Jimmy at School Age," *Child Develop.*, 1939, *10*, 1-19.

CHAPTER 19

Abilities in Later Life

INTELLIGENCE OF ADULTS

STIMULATED by the wholesale testing of army recruits during the first world war, there has been much discussion about the age at which mental growth ceases. The finding that the average recruit possessed a mental age of fourteen years shocked some psychologists into belief that at this age intelligence generally reaches its maximum.[1] Later researches have shown, however, that in certain sectors of the population intelligence continues to grow beyond the age of fourteen. It has been shown also that in later adulthood there is generally some decline in mental ability.

Changes with age. The relationship between the intelligence test score and chronological age for persons of various ages was studied by Miles and Miles.[2] In a city of 15,000 population, the Otis self-administering test of intelligence was given individually to 701 persons 7 to 94 years of age, 616 of whom were 15 or older.

During the ages from 7 to 17 years inclusive there was a gradual rise in score (raw score, not IQ) with age, as indicated by a correlation of $.80 \pm .02$ between score and age. At the age of 18 years the maximum scores in this population were obtained, followed by a plateau in the twenties and a later persistent decline. The decline in ability from the persons who

[1] See discussion on this point by Frank N. Freeman on pages 357-364 of *Mental Tests: Their History, Principles and Applications*. Boston, Houghton Mifflin Co., 1926, pp. ix, 503. Freeman pointed out certain paradoxes and difficulties in accepting this view.

[2] Miles, Catharine Cox, and Miles, Walter R., "The Correlation of Intelligence Scores and Chronological Age from Early to Late Maturity," *Amer. J. Psychol.*, 1932, *44*, 44-78.

were 15 years of age to those who were 95 years of age is indicated by a negative correlation, —.53 ± .02.

When the scores were expressed in terms of IQ, it was found that the population studied was somewhat superior in early adulthood, having a mean IQ of 114 in the teens and twenties (ages 15 to 29 years). Persons 70 to 89 years of age had a mean IQ of 88, or 26 IQ points lower than the group between 15 and 29 years of age. The decline with age was evident for both men and women.

TABLE 20

	Both		*Males*		*Females*	
Age, years	*Number*	*IQ*	*Number*	*IQ*	*Number*	*IQ*
15 to 29	131	114	64	117	67	110
30 to 49	177	108	79	107	98	109
50 to 69	222	99	93	103	129	97
70 to 89	86	88	36	90	50	88

Relation to formal education. Groups who had different amounts of formal education differed in IQ at each age. Each group decreased in IQ with age in about the same amount. The persons who had only grade school education decreased to low dull-normal and borderline between dull-normal and feeblemindedness in the seventies and eighties. Those who had some college training were superior in intelligence in the twenties and average in ability in the seventies and eighties.

TABLE 21

	Grade school (0-8 yrs.)		*High school (9-12 yrs.)*		*College (1-10 yrs.)*	
Age, years	*Number*	*IQ*	*Number*	*IQ*	*Number*	*IQ*
20 to 29	3	101	24	107	51	118
30 to 39	11	94	36	106	39	116
40 to 49	18	93	29	105	37	117
50 to 59	41	89	62	100	42	111
60 to 69	45	85	51	95	35	106
70 to 79	36	82	30	95	19	100
80 to 89	12	75	6	85	1	91
90 to 99			1	79		

Retests of college students. Studies of college students who have survived to the senior year indicate that mental growth

for them does not reach its maximum even at 18 years, but continues throughout the period of enrollment in college. Recently, several investigators have reported retests in the senior year of students who were first tested as freshmen. All studies report an increase in intelligence test score; each of the six investigators [8] whose results are shown here found that the increase was statistically significant. In one study, that made by Rogers at Vassar, the freshmen students were very high in score and made only a small increase, which, however, was statistically significant.

In Livesay's study every individual gained in total score from the freshman to the senior year; in McConnell's study 4 students made lower scores as seniors than they had as freshmen; in Flory's study 12 students showed losses, 5 showed no change in score, and 57 gained.

TABLE 22

Investigator	*Test*	*Number*	*Freshman score*	*Senior score*	*Change in score*
McConnell	American Council on Education	70	144.8	185.2	40.4
Livesay	American Council on Education	50	154.7	199.5	44.8
Flory	American Council on Education	74	72.7 *	81.3 *	8.6 *
Rogers	Thorndike	56	88		5.4
Wolcott	Thorndike	190			60.5
Hartson	Ohio State University	403	73	84	11

* Scores expressed as percentile ranks.

[8] Flory; Charles D., "The Intellectual Growth of College Students," *J. Educ. Res.*, 1940, *33*, 443-451.

Hartson, L. D., "Does College Training Influence Test Intelligence?" *J. Educ. Psychol.*, 1936, *27*, 481-491.

Livesay, T. M., "Does Test Intelligence Increase at the College Level?" *J. Educ. Psychol.*, 1939, *30*, 63-68.

McConnell, T. R., "Change in Scores on the Psychological Examination of the American Council of Education from Freshman to Senior Year," *J. Educ. Psychol.*, 1934, *25*, 66-69.

Rogers, Agnes L., "The Growth of Intelligence at the College Level," *Sch. and Soc.*, 1930, *31*, 693-699.

Wolcott, Willa, "Changes in Thorndike Intelligence Test Scores at the End of the College Course," *Sch. and Soc.*, 1933, *37*, 630-632.

Wolcott made an analysis of changes in score in relation to the degree of concentration of courses in the major department. She noted that there was a wide difference in the practice of major departments in regard to the proportion of course work required. Students who majored in the departments which required the greatest concentration of courses made the least gain, while those students with the greatest variety of courses made the greatest gains. The difference is expressed by a correlation of $-.68 \pm .15$ between amount of concentration in one department and gain in intelligence test scores. At Mills College, where this study was made, the departments with greatest concentration were music, English and art. The departments with least concentration were education, the sciences, and social institutions.

In a related experiment Hartson concluded that "improvement in intellectual abilities continues during the college period in the fields in which that experience provides exercise" (p. 490). He found that on the sub-tests requiring numerical computations a large proportion of the students made higher scores when they were freshmen than when they were seniors. Students majoring in mathematics or science improved in tests employing numerical symbols; in these tests students majoring in English and the language groups made a poorer showing.

MANUAL SPEED OF ADULTS

Studies of manual motility and reaction time throughout the life span indicate an increase in speed through childhood and a decrease in speed in late maturity. Miles [4] has given results on several different skills for persons ranging in age from 6 to 95 years. The age at which maximum speed was attained varied with the different skills, but was usually in the twenties.

Age and manual skills. The scores on speed of manual reaching and grasping presented in Table 23 illustrate the change with age. For testing this skill a special electric clock

[4] Miles, Walter R., "Measures of Certain Human Abilities Throughout the Life Span," *Proc. Nat. Acad. of Sciences*, 1931, *17*, 627-633; "Training, Practice and Mental Longevity," *Science*, 1935, *81*, 79-87.

was used. The clock started when the subject let go the clock key and began to reach; it stopped when he completed the required act and returned to the key. The procedure was as follows: starting from the clock key the subject reached six inches to one side, grasped a short round pencil which was standing in a vertical hole, thrust the pencil into another opening 1½ inches farther away, then returned his hand to the key. The average times required to complete this task are given in seconds for 10 different age groups, totaling 331 males.

TABLE 23

Age of subject, years	*Number*	*Mean score, seconds*
6 to 9	10	1.53
10 to 14	17	1.37
15 to 19	21	1.19
20 to 29	42	1.18
30 to 39	40	1.23
40 to 49	39	1.24
50 to 59	58	1.30
60 to 69	52	1.39
70 to 79	39	1.53
80 to 89	13	1.86

Of the 52 men who were 70 years of age or older, the fastest third averaged 1.35 seconds, a speed of performance equal to the average ability at 60 to 69 years and at 10 to 14 years.

The correlation between age and score for 283 individuals 20 to 89 years of age was —.535 ± .03, indicating a tendency toward decrement in score with age. In view of the moderately low correlation, however, Miles concluded that the score variance in these abilities appeared to depend as much or more on the sum total of other factors as on chronological age.

Summary and conclusions. The curve of mental growth shows an increase in ability throughout childhood and a decrease in ability in later maturity. College students who survive for the full four-year course generally increase in intelligence test scores from the freshman to the senior year.

Manual skills show a somewhat similar pattern of increase in speed through childhood and a decrease in speed in later adulthood.

Part Six

EDUCATIONAL PSYCHOLOGY

By

DAEL L. WOLFLE

University of Chicago

CHAPTER 20

Primary Mental Abilities

PSYCHOLOGISTS have attempted to measure human intelligence in many ways. Some tests have been very simple, such as measures of reaction time or speed of tapping; others, more complex, such as tests of memory, judgment, or reasoning ability. Some tests have been of the paper-and-pencil variety while others have called upon the subjects' ability to manipulate various objects, but the most generally useful of all have been those which call for a variety of fairly complicated mental processes. This is probably so because intelligence tests have most commonly been used to predict success in school, and success in school obviously depends upon a variety of intellectual skills. Tests of this type are frequently called tests of general intelligence.

The concept of general intelligence. Despite its usefulness in school situations, the concept of general intelligence has never been an entirely satisfactory one. Frequently people are found who can do one kind of intellectual work very well while achieving only mediocre success in other lines; conversely, an occasional person is found who is able in most lines but strikingly poor in some special field, suggesting that there may be several different kinds of intellectual ability instead of just one general kind.

Of the various methods used in trying to isolate separate kinds of intelligence, the most effective is the method of factor analysis. A factorial analysis of intellectual ability starts with the assumption that if there are *several* distinct sorts of ability, those tests which involve one type will correlate fairly highly with each other and not so highly with tests which involve a different type of ability. Accordingly, one constructs a large number of different tests, selecting some of one kind and some

of another so that the various types of ability which are thought to exist will all be represented. After these tests have been given to a number of subjects, the correlation between each test and every other individual test is computed. The correlations are then studied by the methods of factor analysis in a search for clusters of tests which correlate highly with each other. The tests making up each cluster are then examined in an attempt to find what they have in common—the *common factor*. If the results are clear enough, the experimenter may give the factor a name which indicates the type of ability which he thinks it involves.

The Thurstone ability tests. One of the largest and most thorough factorial studies of mental ability was conducted by L. L. Thurstone, who started with 57 different tests divided into 10 general types intended to measure the various kinds of ability which he thought might exist.

Of five tests of "the ability to abstract" one consisted of a number of statements or proverbs. In each item of this test the subjects were instructed to check two of the proverbs which had nearly the same meaning as a given statement at the top; for example:

Sail when the wind blows.

—— Strike when the iron is hot.
—— One must howl with the wolves.
—— Make hay while the sun shines.
—— Make not your sail too large for the ship.

Other tests of the ability to abstract demanded that the subject classify similar words, or that he mark simple geometrical forms according to whether they belonged to one or the other of two groups of figures given them as samples. The instructions and some sample exercises from this latter test are shown in Fig. 31.

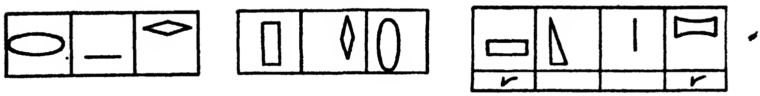
There were eight tests of verbal ability. In one of these the subject had to write as many synonyms as he could think of for each of several words such as *happy* or *large*. In another test he had to write as many words as he could that begin with

FIG. 31. Samples from Figure Classification (28 Items).

In each line below, there is a rule by which the symbols in Group I differ from those in Group II. *There is a new rule for each line. Your problem is to discover the rule in each line.* Some sample problems are worked for you below.

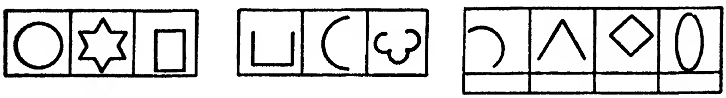
In the first line below the rule is that the symbols in Group I are *horizontal* while those in Group II are *vertical.* Each of the *Test Symbols* at the right belongs either to Group I or to Group II. The test symbols that belong to Group I have been checked.

GROUP I GROUP II TEST SYMBOLS



The rule in the problem below is that the figures in Group I are *closed* while those in Group II are *open.* Now check the test symbols that belong to Group I.

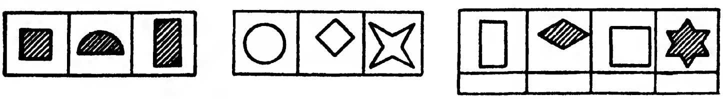
GROUP I GROUP II TEST SYMBOLS



You should have checked the third and fourth symbols. They are closed figures.

Now find the rule in the following line by which Group I differs from Group II. Then check the test symbols that belong to Group I.

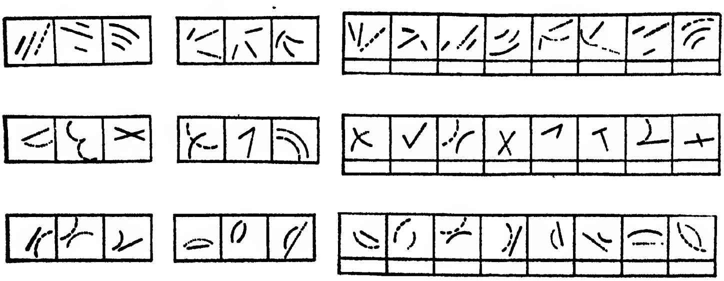
GROUP I GROUP II TEST SYMBOLS



You should have checked the second and fourth symbols. They are shaded.

Now mark the following exercises in the same way.

GROUP I GROUP II TEST SYMBOLS



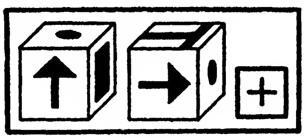
s and end with *l*. In a third verbal test the words of a sentence were given in mixed-up order, and the subject had to indicate whether the sentence would be true or false if it were properly rearranged. The other five tests in this group were all designed to measure some phase of the ability to handle words and sentences.

The third group of tests was intended to measure the ability to think in terms of space or of movement in space. One of these is shown in Fig. 32. Another space test showed a num-

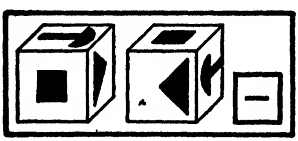
FIG. 32. Samples from Cubes (32 Problems).

The drawings in this test represent cubes. There is a different design on each face of the cube. A cube has *six* faces.

Notice that both of the drawings below can represent the *same* cube. *Be sure you see that the first and second drawings represent the same cube turned into two different positions.* Since both drawings can represent the same cube, a plus sign (+) has been placed in the blank square at the right.

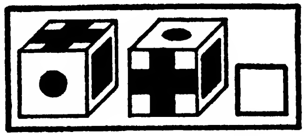


Notice that the two drawings below represent two *different* cubes and a minus sign (—) has been placed in the blank square at the right. *Be sure that you see that it would be impossible* to turn the cube shown in the first drawing so that it would look EXACTLY *like the cube shown in the second drawing.* Unless you see this clearly you can not solve the test items. There is a different design on each face of the cube.



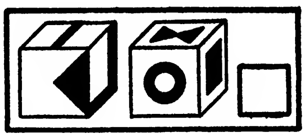
If the next two drawings *can* represent the same cube turned into different positions, put a plus sign (+) in the square at the

right. If the two drawings below *can not* represent the same cube, put a minus sign (—) in the blank at the right. *Remember that there is a different design on each face of the cube.*



You should have marked the above problem minus (—). Study it carefully to be *sure* you see that this is true.

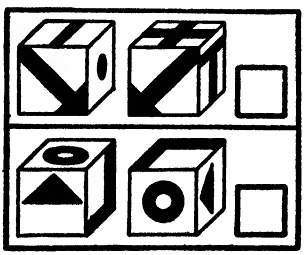
Now mark the problem below.

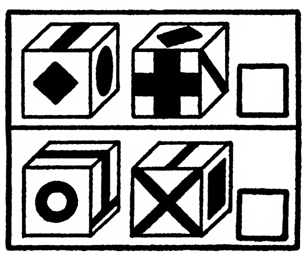


You should have marked the above problem plus (+).

In the following practice problems, put a plus sign (+) if the two drawings can both represent the same cube. Put a minus sign (—) if they can not both represent the same cube. Remember that there is a different design on each of the six faces of a cube.

Go right ahead. Do not wait for any signal.





ber of pictures of connected gears, belts, and levers. The subject was told in which direction one gear was turning. He was then asked to indicate how some other part would move. An example is given in Fig. 33. In another test pairs of pictures of the American flag were shown. Sometimes the two pictures

in a pair represented the same side of the flag; sometimes they represented opposite sides. The subject was instructed to indicate whether the same or opposite sides were shown. In all there were nine different space tests.

Four of the tests measured the ability to recognize similarities and differences in geometrical forms. In one such test a figure was shown followed by five other figures, one being exactly like the first figure, the other four similar to the first one but not identical with it. The identical one was to be picked out. Part of this test is reproduced in Fig. 34.

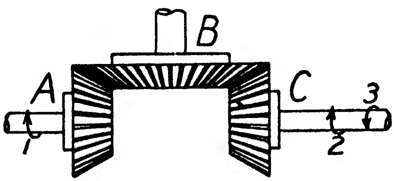


FIG. 33.

A, *B* and *C* are three beveled gears.

1. If *A* is turning in the direction shown, which way is *C* turning, 2 or 3? . . . ———.
2. Can B be the driver? . . . Yes No.

(Draw a ring around the correct answer.)

Of the six tests of arithmetic four consisted of relatively easy addition, multiplication, division, and subtraction. One required translating a code consisting of dots and bars into ordinary numerals and solving problems given in the code. The other numerical test consisted of a table with some gaps in it which the subject could fill in from the data given in the rest of the table.

Four tests under the heading "Numerical Reasoning" were included. One required the subject to fill in the two missing numbers in series such as 27, —, 23, 23, 19, 19, —. Another test of numerical reasoning was made up of problems with four answers given for each. The one right answer was to be marked, but the subject did not have to work through all of the arithme-

tic to determine which was correct. For example, in one problem 4.12395 was multiplied by 6.82187. The four answers were 7.563327, 14.012468, 28.133051, and 56.103378. The correct answer is obviously approximately 28, so the third answer can be marked immediately. Subjects could save a good deal of time by not working out the problems completely but only enough to find the one right answer out of the four alternatives.

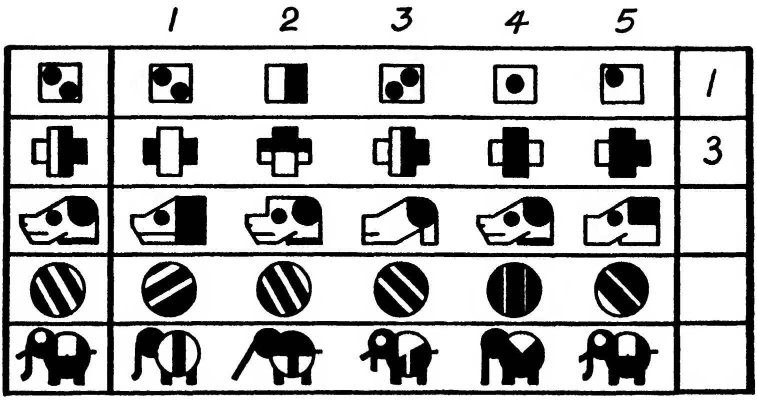


FIG. 34. Samples from Identical Forms (60 Items).

The first figure in each line below is exactly the same as one of the five numbered figures following. In the blank space at the right of each line, write the number of the figure which is exactly the same as the first figure in the line. The first two blank spaces have been filled in correctly. You fill in the other three. Go right ahead. Do not wait for any signal.

Three tests of verbal reasoning and three of space reasoning were included. One of the verbal reasoning tests consisted of nonsense syllogisms which were to be marked in one way if the conclusion was justified and in another way if it was not. One of the syllogisms was:

All haystacks are catfish.
All catfish are typewriters.
Therefore all haystacks are typewriters.

In this example the conclusion is logically correct even though the premises are both absurd.[1] One of the space reasoning tests was made up of form analogies. In one example, where a large

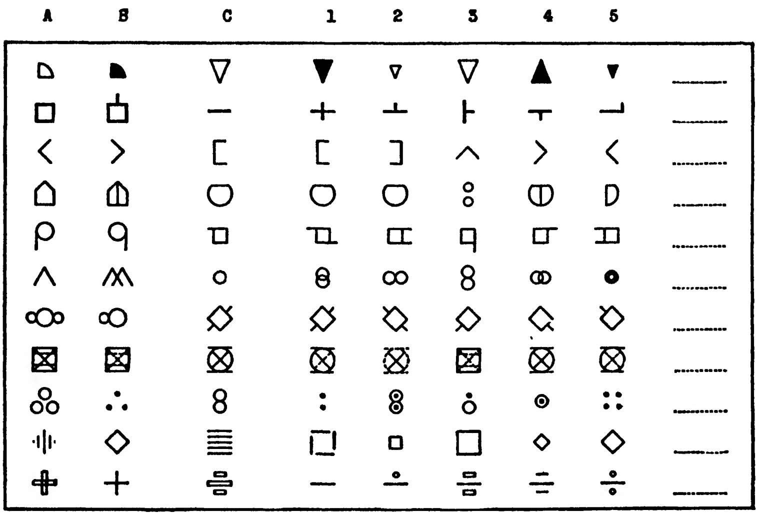


FIG. 35. Sample Items from Pattern Analogies.

First, decide what rule is used to make figure B from figure A.

Second, apply this rule to figure C and find the resulting figure among the figures 1-5.

Third, write the number of this figure in the blank at the end of the row.

Go ahead. Do not wait for any signal.

circle and a small circle were followed by a large square, the subject had to select that one of five other figures (a small square) which bore the same relation to the large square as the small circle did to the large one. Some of the test problems are to be found in Fig. 35.

[1] If the conclusion does not seem justified, try substituting letters for the terms in the premises. The syllogism then assumes the familiar form

All A's are B
All B's are C
Therefore all A's are C,

and the logical correctness of the conclusion becomes apparent.

There were six tests of rote learning. In each test the subjects were instructed to learn a list of words, pairs of numbers, names and initials, or other material. Appropriate recall tests were then given. In one, the subjects were asked to write the initials when only the last names were given to them. In some of the recall tests a long list was given and the subjects were asked to mark those items which were in the list which they had just learned.

Finally there were nine miscellaneous tests which did not fit very well into any of the above groups. Two measured the size of the subject's vocabulary. One was a spelling test and one a grammar test. Another demanded that the subject write a short theme describing some person whom he knew well. Two scores were given on this test. The themes were graded by English instructors, and the total number of words written was counted. The first score indicated quality of work; the second, it was hoped, might measure some aspect of speed of thinking or mental fluency. Still another of these miscellaneous tests showed a number of pictures of human hands in different positions. The subject had to indicate whether each picture represented a right hand or a left hand. The other two tests were intended to measure auditory factors and the sense of rhythm.

All 57 tests were given to 240 university students who volunteered to serve as subjects. About 15 hours was required for the tests. The subjects were well motivated by their interest in the study and by the promise of an analysis of their own intellectual strong points and shortcomings.

Factor analysis of tests. The 57 sets of test scores were then correlated with each other and the correlations studied by the method of factor analysis to isolate the factors involved in the original tests. The statistical part of a factor analysis ends with a table of factor loadings from which one can read the weight, or importance, of each factor in determining the score on each test. Table 24 shows the weights of each factor (the factors are interpreted below) in each of several of the tests which have been described above.

When the statistical analysis is complete, it is necessary to

examine the table and pick out those tests which involve each of the factors. By comparing tests which show high weights on one factor with tests which show low or zero weights on that factor it is possible to get some ideas as to the nature of the factor involved. Thurstone's examination of the table of factor weights computed from the intercorrelations of his 57 tests showed several distinct and consistent factors.

One factor appeared prominently in all of the tests of a spatial nature, both those previously classified as "space" tests and those designated "form" tests. It also appeared in some miscellaneous tests, as the one requiring the subject to note whether each of a number of pictures represented the right or the left

TABLE 24

SAMPLE OF THE FACTOR LOADINGS FOUND IN THURSTONE'S ANALYSIS OF 57 TESTS

	Loadings on factor							
Name of test	*S*	*P*	*N*	*V*	*M*	*W*	*I*	*D*
Reading (proverbs)	.002	.169	—.130	*.552*	.208	—.007	.306	.148
Figure Classification	*.393*	.196	.080	.054	—.052	.058	*.405*	*.398*
Inventive Opposites	—.016	.142	.003	*.450*	.160	.072	.070	—.041
Cubes	*.626*	.211	.199	—.065	—.034	.145	.099	.272

The four tests are described in the text. The initials S, P, N, V, M, W, I, and D refer to the eight factors: space, perception, number, verbal, memory, word, induction, and deduction. These factors are discussed in the text.

The table is to be read as follows: The reading test had a loading of .002 on the space factor, a loading of .169 on the perceptual factor, a loading of —.130 on the number factor, and the other loadings as given in the body of the table. Each of the other lines may be interpreted in a similar manner.

In order to be of much use in determining the nature of the factor in question, a test loading should be about .40 or higher. Thus, in trying to interpret factor S, the experimenter would examine tests which have loadings of .40 or higher, the Figure classification and Cubes tests and all of the others not shown in the table which have such loadings, and contrast them with those tests such as Reading and Inventive opposites which have low loadings. By comparing these two sets of tests he may be able to discover what it is that the tests with high loadings have which those with low loadings lack.

The high loadings which are useful in determining the nature of the factors, and which are useful in describing the factorial composition of a particular test, are printed in italics. Note that each test has high loadings on one or more of the factors, and low or nearly zero loadings on most of them. The interpretation given to this distribution of factor loadings is somewhat as follows: The Reading test, as an example, involves the verbal factor quite prominently and the induction factor to a smaller extent. The other factors are not involved to any important extent in this test.

hand. Since the factor appeared in the space and form tests regularly and with high weights, and did not have high weights in the non-space tests, it has been called the *space* factor.

The second factor had large weights in nine tests. Some of these were verbal in nature, others were not. One of the nine tests was the one which required the subject to select quickly that one of a group of geometrical figures which was identical with another given figure. The important common element in all nine tests seemed to be the ability to pick out some detail quickly when that detail is buried in irrelevant material. To indicate this characteristic, Thurstone named this the *perceptual* factor.

The third factor appeared prominently in the arithmetic tests, but not in the tests which had been intended to measure arithmetic reasoning. It was most important in the simpler tests and had less weight in the more complex ones. It was called the *number* factor. It seems to be the ability to deal rapidly and accurately with simple numerical operations.

The fourth factor had highest weights in those tests which measured reading ability (for example, the proverbs test illustrated above) and in those which required the subject to write down as many synonyms or opposites as possible for a given word. It appeared to be an important factor in all of the tests which required some kind of verbal reasoning or required the subject to deal with the meanings of words. Thurstone named it *verbal relations* or the *verbal* factor.

The fifth factor also involved a verbal ability but one of a quite different sort. Where the verbal factor emphasized the meaning and logical relations of words, this factor seemed to involve the mechanics of single words. It had high weights on those tests in which the subject had to write as many words as he could that began and ended with designated letters. It also had high weights on the spelling and vocabulary tests, and on tests of unscrambling disarranged letters to make a word. Since single words rather than verbal relations appeared to be the important element in this factor, Thurstone named it the *word* factor.

The sixth factor had its highest weights in the rote-learning tests. It was named the *memory* factor.

The seventh factor had high weights on five tests, some of which were numerical and some of which were spatial. In all five it was necessary for the subject to pick out some rule or principle which bound the items together. This common element is indicated by naming it the *induction* factor.

The last factor was more difficult to identify because it had high loadings on only four tests. Since these all involved some kind of deductive reasoning, the factor has been named the *deduction* factor.

The factor analysis condensed the original 57 tests into eight factors: space, perception, number, verbal relations, word, memory, induction, and deduction. In connection with these findings three questions should be raised: How consistently do these particular factors appear in other studies? How do they differ from the several types of ability which Thurstone assumed at the beginning? How can knowledge of the factors be used?

The fact that the same factors have reappeared when Thurstone has analyzed new batteries of tests increases greatly one's confidence in their importance and generality. This feeling is further strengthened by finding the same factors appearing in batteries of tests studied by other investigators. Those factors about which one can be most certain are number, verbal relations, space, and memory. The perceptual factor has been reported by several other workers. The induction and deduction factors are less constant; at present it is not certain just what their natures may be. The word factor has failed completely to appear in some studies and has been clearly present in others. The fact that a particular factor fails to appear in one factor analysis does not mean that its appearance in another study was an accident. No factor will show up unless several tests that involve it are included in the battery. The non-appearance of the word factor may merely mean that the investigator did not include tests which involve it.

Some of these factors would probably be included if one

were asked to list the different types of mental ability without knowing the results of factor studies. Verbal, number, and memory factors are examples. But not all of the factors are of this type. Thurstone expected to find a verbal factor; he found not one but two factors prominent in his verbal tests. He expected to find a factor of ability to think in terms of solid space and another of ability to think in terms of flat space; instead he found only one space factor. The number factor was more restricted than he had expected it to be. It seems to involve only the relatively simple numerical operations. More difficult mathematical problems involve this factor to a very slight extent and depend much more on the reasoning factors. In short, the factors show some similarities to one's preliminary guesses. But factor analysis also reveals a number of unsuspected relationships, and indicates just which tests involve each of the factors much more clearly than one can tell in advance.

With these factors isolated, it has become possible to construct intelligence tests which measure each of them separately. These new tests may replace the general intelligence tests which give an average of all of the factors. But for many purposes such an average is highly useful. Since all of the factors are involved in highly complicated intellectual tasks such as getting a college education, a general measure of intelligence predicts one's aptitude for college or university work fairly well. But if one is interested in predicting the probable success in a particular course or a particular line of work, measuring the factors separately may provide even better prediction. As a start in this direction, Thurstone has prepared a battery of 16 tests which provide perceptual, number, verbal, space, memory, induction, and deduction scores. When these 16 tests are taken, the scores can be presented in the form of a profile which readily reveals the abilities in which a person makes high scores and those in which his scores are low. Such a profile is shown in Fig. 36.

Thurstone writes:

It is too early in these investigations to make any definite statements about the particular combinations of abilities that are called for by each vocation or to attempt to make individual vocational prognoses by these scores. Some of the combinations seem rather obvious. For example, one might expect an accountant to have a high rating in the factor N. A law student might have a high score

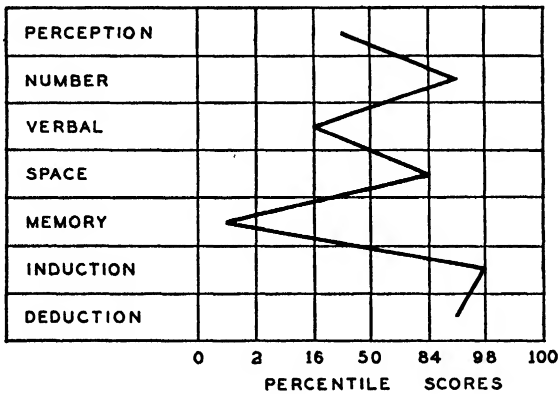


FIG. 36. Ability profile of one individual. The seven abilities are named at the left. The percentile scores indicate the percentage of a normal group of people whose scores are lower than those of this individual. The profile indicates that the score made on the perception test was slightly below average. That on the number test was better than some ninety per cent of the similar scores. The other points can be interpreted in a similar manner.

in V relative to his other scores. An engineering student would probably have a high rating in the factor S. However, it is well to recall that a man might be a success in accounting without being quick in adding a column of figures, that he might be a success in some phase of engineering or architecture without being superior in the factor S which may be involved in design, and that he might be a success in law without being superior in the verbal factor V.

With the continuation of investigations we shall be able, eventually, to tell each student much more about his abilities. We shall know more than we do now about the nature of each factor, to what extent it is due to inheritance, to what extent it can be trained, how some factors may serve vicariously for other factors and what factors are involved in different vocations.

The continuing work of Thurstone and the other factor analysts is directed toward answering some of these questions.

References

Thurstone, L. L., *The Vectors of Mind.* Chicago, Univ. of Chicago Press, 1935.

A detailed description of the mathematical techniques which Thurstone uses in factor analysis.

Thurstone, L. L., *Primary Mental Abilities.* Psychometric Monographs, 1938, 1, No. 1.

The complete account of the work reviewed in this chapter.

Wolfle, D. L., *Factor Analysis to 1940.* Psychometric Monographs, 1940, 1, No. 3.

A review of the development of factor analysis, of its chief problems, and of the principal findings.

Learning to Read

NEARLY everyone learns to read, and continues after school days are over to read newspapers, magazines, or books for study, information, and fun. Since reading is so widely practiced out of school, and since work in all the other school subjects is so dependent on this ability, reading is probably the most important skill youngsters learn in their early school careers. Because of its importance, psychologists, educators, and school teachers have spent a great deal of time [1] finding out how people read, how good readers differ from bad ones, how children can best be taught to read, and how poor readers can improve. The combined results of all this work have made the teaching of reading a very different matter from what it was when Noah Webster wrote his famous *Speller* in 1783.

The A-B-C method. Webster started systematically with the alphabet and the sounds of the English language. When children had learned their A-B-C's, they progressed to simple combinations of a consonant and a vowel, and learned *ba, be, bi, bo, bu, by, ca, ce, ci, co,* and so on until they knew all of the two-letter combinations. After these came simple two-letter words in easy combinations, such as *go on, go in, go up, an ox, by me.* Later stages introduced three, four, then five letter words. When their training in reading and spelling was ended—and it continued throughout the school years—many of the pupils were still such poor readers that Horace Mann reported in 1838: "More than eleven-twelfths of all the children in the reading-classes, in our schools, do not understand the meaning of the words they read; . . . the ideas and feelings intended by the author to be conveyed to, and excited in, the reader's

[1] Over 1200 studies of reading were published between 1925 and 1937

mind, still rest in the author's intention, never having yet reached the place of their destination."

A-B-C method criticized. Others besides Horace Mann became dissatisfied with the shortcomings of the A-B-C method of teaching reading and suggested improvements which might make learning to read an easier task. Some systems started by having the children read whole words and phrases without analyzing them into letters; others employed phonetic methods of breaking words into syllables or sound units. Some trained the beginners with families of words that ended alike, such as *sit, bit, pit, hit, mit,* while still others thought it better to emphasize groups which *started* with similar sounds, such as *cat, can, cab, calf, cap.* Sometimes these systems had little more to recommend them than the enthusiasm of their authors, but they all had in common a recognition of the failure of the earlier methods, and they all indicated the hope that better methods might be found. Out of this dissatisfaction grew the truly experimental attacks on the problem.

The mechanics of reading. The basic problem of teaching children to read can be stated very simply. One must learn how good mature readers read; one must find out how beginners read; one must then find the most efficient methods of converting the beginner into the accomplished mature reader. The answers are harder to find than the questions.

One of the easiest aspects of the problem to start working on is the sheer mechanics of reading—what one's eyes do in reading. In the latter part of the nineteenth century it was discovered that the eyes move in a series of jerks across the lines of print. The eyes fixate one point near the beginning of the line, move quickly to another point, fixate it for an instant, move rapidly again, and so on. Only 6 to 10 per cent of the time is spent in these eye movements; the rest is taken up by the fixation pauses. It is during these pauses that the actual reading is done.[2]

[2] Woodworth, R. S., *Experimental Psychology,* pp. 591-594, describes several simple and interesting demonstrations of how little can be seen while the eyes are in motion.

Cameras were soon developed to photograph these eye movements. A schematic drawing of such a camera is shown in Fig. 37.

Eye-movement cameras photograph a beam of light reflected from the cornea of the eye. When the eye moves, the angle of reflection of the light will be changed and the position of the light track on the moving film will alter accordingly. By interrupting the beam of light with a large vibrating tuning fork, the photographic image can be changed from a solid line to a series of dots. With the film moving at a constant speed, the beam of light interrupted at known intervals, a record of what was read, and, when the subject is reading aloud, a dictaphone record of the reading, it is possible to reproduce the subject's performance quite exactly.

This method of studying reading was used by G. T. Buswell, who recorded the eye movements of 179 subjects while they read a short simple prose passage. The subjects were fairly evenly scattered in age from first-grade youngsters to university students. Because Buswell wanted to study the normal improvement in reading, he weeded out the best and the poorest readers in each grade and used only the more nearly average ones.

From the photographic record it was possible for Buswell to determine where the eyes were fixated in each line, and for how long a period they stopped at each fixation point. This is illustrated in Fig. 38, in which each vertical line shows the location of one fixation pause. The number at the top of each of these lines indicates whether it was first, second, third, etc., of the fixations in reading a particular line of type. The figure at the bottom of each vertical line gives the length of the pause in units of one twenty-fifth of a second. Fig. 38 presents the record of a first-grade child reading silently, showing that his eyes fixated just to the right of the initial *A* for 22/25 of a second, moving then to the *y* of *boy* for 7/25 of a second, and to the *a* of *had* for 8/25 of a second. The next movement of the eyes, a regression, was backward to the space between *boy* and *had*.

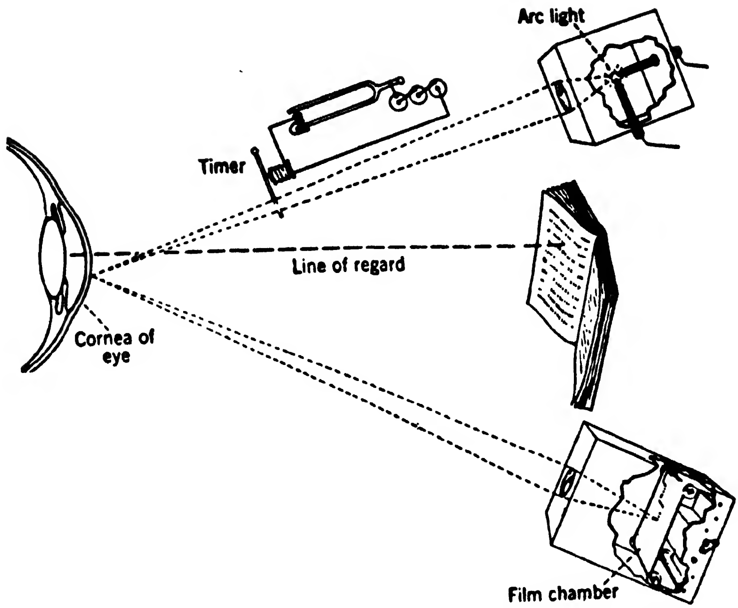


FIG. 37. A method of recording eye-movements in reading.

An arc light projects a light beam through a blue lens that filters out all but the actinic (chemically active) rays (to avoid stimulating the subject). These are projected upon the cornea of the eye, and thence reflected through a focusing lens upon a moving photographic film housed in a dark chamber. Right-left movements of the eyeball produce right-left deflections of the light beam on the film. A signal marker operated by an electrically maintained tuning fork (50-cycle) interrupts the light beam so that it is registered upon the film as a series of dots each of the time value of 1/50 sec. On the moving film eye-fixations show as a vertical row of dots, eye-movements as dashes to right or left. In the figure is suggested a record representing three fixations and two rightward movements.

The film records can be related to the printed matter as follows. The reader first fixates in turn upon a dot placed just above the first letter in the first line and another just above the last letter, producing two guide marks on the film. When finished, the whole film record can be enlarged to the precise dimension of the printed line, and the successive eye-fixations accurately located thereon. (Dodge technique modified by Schmidt, Gray, and others in Judd's laboratory.)

Figure 39 shows a similar record, but one in which the number of fixations was very much smaller. The subject in this case was a college senior. A comparison of Fig. 38 and Fig. 39

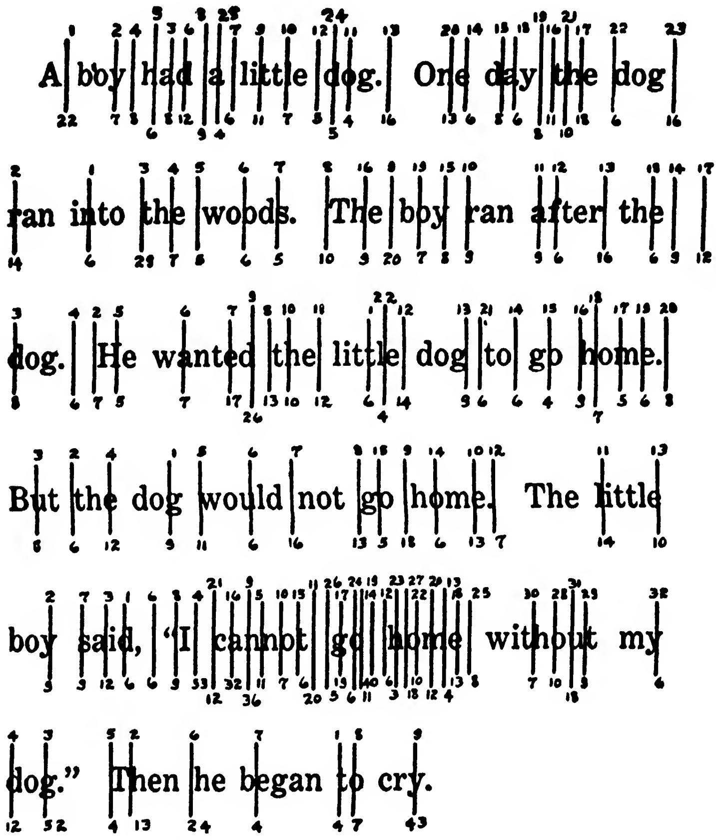


FIG. 38. Silent reading by a first-grade child.

makes some of the differences between beginning and accomplished readers very clear. For example, in the second line of print, the one in which the first-grade youngster showed his best performance, there were 19 fixation pauses as compared to four for the college student. In this particular case the average duration of fixations was not greatly different for the two; 10/25 of a second for the child and a little over 9/25 of a

second for the adult, but the first-grader made five regressions in the line, while the college student made none.

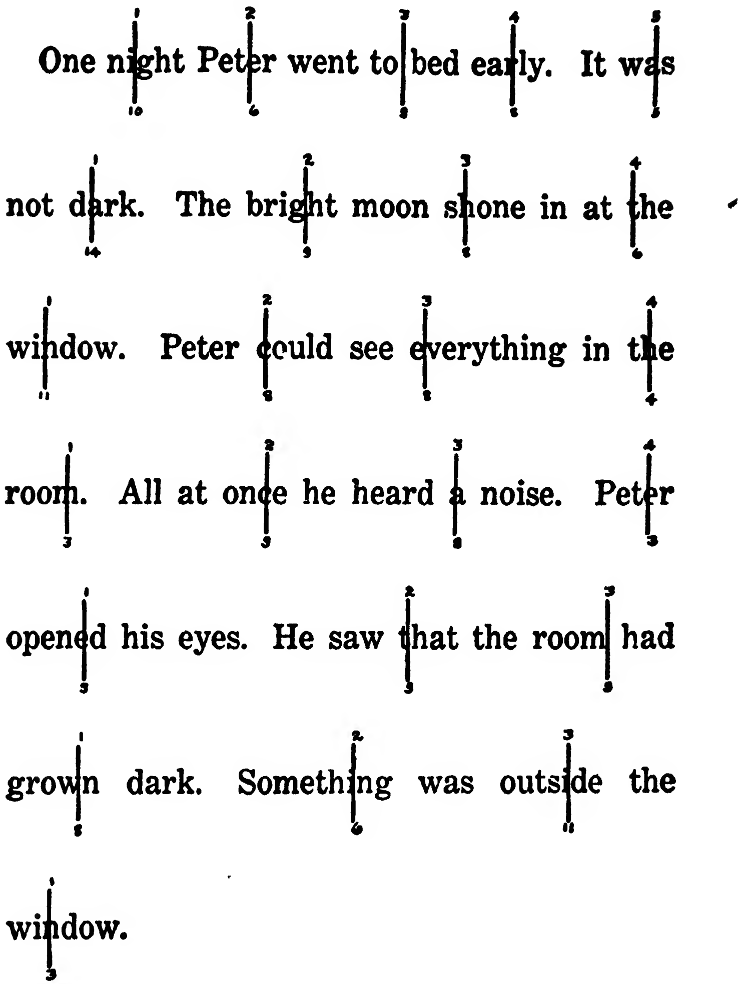


FIG. 39. Silent reading by a college senior.

Differences in reading speed. The difference in speed of reading shown in these two records is obvious. The first-grade child required about eight seconds where the college student required scarcely more than half a second to read a line of print.

The difference in smoothness and regularity of reading is also apparent. For the college student the number of fixations and their duration were fairly uniform from one line to another. For the first-grade reader the number of fixations varied from 13 to 32 and the duration of each varied from 4/25 of a second to more than two seconds.

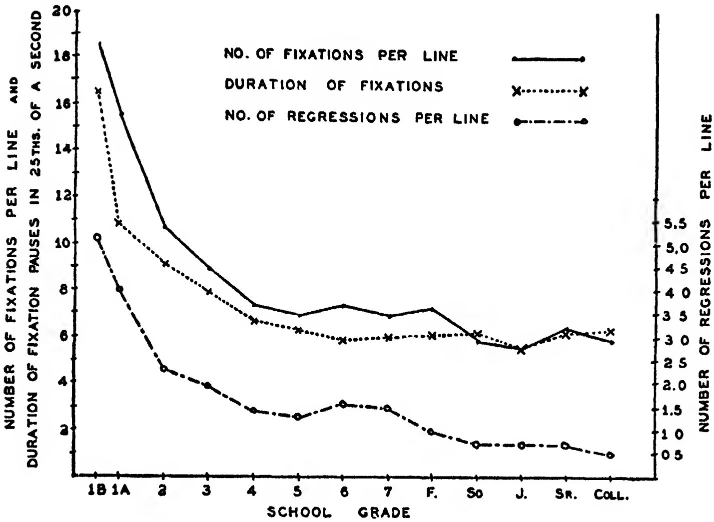


Fig. 40. Improvement in reading. The average number of fixations per line (solid line), the average duration of fixation pauses (dotted line), and the average number of regressive movements per line (dot and dash line) are shown for the different grades from IB through college.

When the three measures—number of fixations per line, duration of fixation pauses, and number of regressive movements per line—are averaged for all subjects within each school grade, the averages may be plotted as they are in Fig. 40. The three curves show the normal improvement as pupils at progressively higher and higher grades are examined. The number of fixations per line drops rapidly from over 18 in the first half of the first year to about 7 by the time the fourth grade is reached. From then on, much less improvement takes place;

college students average six fixations per line. The duration of the pauses also drops rapidly from a starting average of more than 16/25 of a second to about 6/25 of a second in the sixth grade. From that point on no further improvement was shown by Buswell's subjects. The average number of regressions per line was over five for the beginners, dropping gradually to about 1 per line in the upper grades and to about one half per line for high school and college students.

These curves and results from other studies agree in showing the first four or five elementary grades to be the time of most rapid improvement in the mechanics of reading. By the time the sixth grade is reached the reading habits of the children are nearly mature. Some small gain is shown in the curves for the later years, but it is hard to determine what produced that gain. Since Buswell did not study the same children through several grades, but studied different children in the different grades, and since it is known that the less able children are more likely to drop out of school, the improvement from grade to grade might have been due to finding more and more rigidly selected groups as he examined the higher and higher grades. However, Buswell's technique of examining only the average students in each grade tended to exclude from the study the poorest children in the early grades, and since these are the ones most likely to drop out of school, this selection factor is not as important as it would have been had he examined all children in each grade. Nevertheless, selection of subjects from different grades may not give exactly the same results as following the same children year after year from their first attempts to read to their performance in college.

It should be understood that perceptual span (the amount read in one fixation pause) and number of fixations depend upon the material read and would be changed in reading different material. If the lines are shortened or lengthened the number of fixations per line will decrease or increase correspondingly. If a very difficult text is used in place of the easy passage of the experiment, the number of fixations will also increase. Thus, while the three types of record measure directly

the simple mechanics of reading, they also measure indirectly the reading comprehension. In each fixation pause the eyes see part of the line of print. How much will be seen depends upon the ease or difficulty of the material and the perceptual span of the reader. In the last three lines of Fig. 39 it appears that the reader, in his seven fixations, saw seven blocks of material:

opened his eyes.
He saw that
the room had
grown dark.
Something was
outside the
window.

In one fixation only one word was seen, in some two, and in others three words were seen at a time.

Speed and comprehension related. Sometimes a student is found whose mechanical reading skill has outstripped his ability to understand what he reads, but usually speed of reading is no faster than the speed at which one can comprehend. Efficient eye movements, in fact, are not so much the cause of efficient reading as they are an effect or indicator of good reading ability. If one is able to perceive several words at a single glance he will need only a few fixations per line, but if he can understand a word only by laboriously spelling it out, many fixations will be required. Familiar words are easily identified and require few eye fixations; little known words are unfamiliar and require many more fixations.[3] For efficient reading it is not enough to make the eyes move smoothly and rhythmically across the page. It is also necessary to build up a wide vocabulary and the ability to understand and interpret what

[3] Woodworth, *op. cit.*, pp. 737 to 744, summarizes the experimental evidence regarding the important factors in speedy recognition of words. His summary of the findings "allows some cue value to the general shape of the word, especially in indirect vision before the eye reaches the word in continuous reading, but lays most emphasis on the clear view of letters and familiar letter groups, obtained for an instant during the direct fixation of the word" (p. 744).

is read. Various techniques have been used for making certain that growth in comprehension parallels growth in speed. One method, for example, is to test the student's knowledge of the material he has read by asking him questions about it. Such techniques as this are used even when the primary purpose of the reading lesson is to increase speed.

Role of eye movements in reading. Knowledge of the role of eye movements in reading has had a number of results. For one thing it has made possible practical studies of the most effective methods of presenting material to be read, and has suggested studies on the role of eye movements in other fields. Since eye movements are useful indicators of reading comprehension, studies of the best length of line, space between lines, and size and type of letter are useful. Bean, in a study of the eye movements of professional and amateur musicians, found that the average number of notes recognized in a single fixation was nearly twice as great for the professionals as for the amateurs, although some of the best amateurs had larger perceptual spans than two of the professionals. By special training, Bean was able to teach some of the poorer readers to receive patterns of notes instead of single notes, and thus greatly improved their ability to read music. From 20 to 30 hours of special practice produced a substantial improvement in most subjects.

With the widening range of usefulness of eye-movement records (ordinary reading, reading music, reading advertisements, reading chemical and mathematical formulas) have also come improved recording techniques. Clark has described a method of recording vertical and horizontal movements of both eyes simultaneously. Tiffin has devised an ingenious technique for getting simultaneous photographs of the eye movements and of the material being read. An ordinary 16 mm. movie camera is mounted above and about one foot in front of the subject's head. The camera is pointed downward at a half-silvered mirror set at an angle of 45°. One half of the film shows the subject's eyes; the other half shows the page he is reading.

Silent and oral reading compared. Knowledge of the mechanics of reading and knowledge of methods of increasing

comprehension in reading may be applied practically in the elementary school to improve the original teaching, or, in high school and college, to remedial work with students who have never learned to read efficiently. One change in teaching methods has grown out of the finding that silent reading is faster than oral reading, except during the first two or three years of school. Most teaching used to emphasize oral reading. But practically all the reading one does, in or out of school, is silent. Consequently the emphasis has shifted and most of the teaching of reading now aims at making rapid silent readers rather than polished oral readers.

Differences in reading ability. Another result which has been found time and again in studies of reading is that at any age, in any grade, and at any level of mental ability there are wide differences in reading ability. Gumbinger has made this finding the starting point in a study of the relative effectiveness of three methods of improving reading achievement in the sixth and seventh grades. Her three groups of students were so selected as to be equal in mental ability and in reading ability at the beginning of the experimental period. One of the groups was taught reading by the "unit-mastery" system. No class work was given to this group. Instead, each student was given a reading lesson specifically designed for his own level of ability. Each lesson consisted of a story to be read and a series of questions to be answered. When the student had correctly answered all of the questions on one lesson he turned it in and was given the next one. This method allowed the students to start at different levels of ability and to progress at different rates. Each student was always working on material designed for a child of his particular level of achievement.

The second group of students was taught by the same "unit-mastery" method for half of the time, and by the usual textbook method the other half of the time. This group had some of the advantages of the individual method, if any were present, and some of the advantages which might be inherent in the group method of teaching.

The third group was taught by the usual textbook method

However, attempts were made to individualize some of the training even in this group. One day of each week was set aside for remedial work with the poorer readers. This work was intended to correct the particular difficulties of any child who was not getting along well. In addition, the abler readers were encouraged to read more than the less able ones.

While each of the three groups had some individual work, the amount of individualization of instruction definitely decreased from the first to the third group.

The experiment lasted for a school year. At the end of the year the members of all three groups were again tested to determine their reading ability. The tests given were intended to measure several aspects of reading ability: meaning of individual words, paragraph meaning, rate of reading, knowledge of facts read, knowledge of the central thought of a passage, or of the passage as a whole, ability to understand and follow directions, and ability to organize the material read. Each child was scored in terms of his improvement during the year. The scores were expressed in terms of the number of "years of improvement," where one year of improvement is the gain expected for the average pupil in a year of ordinary school work.

The average gain for the first group (taught by the unit-mastery or completely individualized method) was one year, three and one-half months. The average gain for the second group (taught by the half and half method) was one year and four months. The average gain for the third group (taught by the usual textbook method) was one year and one month. All three groups gained more than is normally expected in a year of schooling, but the two groups which had emphasis upon individual work gained more than the group which devoted most of the time to regular class instruction. Though none of the differences between averages of the three groups satisfied the criterion of statistical reliability, most of the results on the different tests were consistent in pointing to the inferiority of the method used with the third group.

When the data were analyzed in more detail in an attempt to determine the relative effectiveness of the three methods for

students of different levels of ability, it was found that the best students gained much more by the unit method than by the textbook method. The average students gained somewhat more by the unit method, and the poorer students gained only a little more by the unit method than by the text method.

The author concluded that the individualized method was superior to the group method in teaching reading and that the difference in effectiveness of the methods was greater for the better students than for the poorer ones. There was enough overlapping of the groups at the end of the experiment so that these conclusions could not be made with absolute certainty on the basis of this study alone; however, results from other studies support them.

In a study such as this there are always a number of questions left quite unanswered. For example, the unit method was new to those students who used it; how much of its apparent superiority was due to the teaching method and how much was due to its novelty? Would the textbook method be found superior if it were introduced as an experiment into a school which regularly used the unit method?

Another question: The same teacher (the experimenter) taught all three classes; how much do the results of the study depend upon the ability and enthusiasm of the teacher? Did she work particularly hard with one group and unintentionally stir up enthusiasm in that group more than in the others?

If one is interested in the comparison of the group and individual methods of teaching in this study, it is important to ask about the nature of the individual work that was given to the members of the third group. Did the teacher bring about more improvement in this group than other teachers would have secured simply because she was a reading expert and worked so effectively with the children individually one day each week that the deficiencies of the group method did not show up as obviously as they might have under another teacher?

These questions are raised to emphasize the fact that caution is necessary in interpreting the results of studies of this type.

Results in the first and third groups were certainly different. Statistical computations can indicate the degree of reliability of that difference, but they cannot give any information on the cause of the difference. It might conceivably have been due to differences in ability. It probably was not, for the experimenter matched her groups to start with. It might have been due to differences in motivation and enthusiasm of the pupils in the different groups, or it might have been due to differences in the effectiveness of the methods of teaching. By basing conclusions on a number of experiments or by carefully setting up experimental procedures in such a way that some of these questions can be definitely answered, it is possible to arrive at more positive conclusions than can be reached on the basis of this experiment alone.

The experimental studies which have been reviewed in this chapter represent a very much larger group of studies. Many of them, like Buswell's, were designed to learn more about the processes involved in reading; many, like Gumbinger's, were intended to test out the usefulness of methods of teaching based on previous experimental findings. The combined results of all these studies justify the statement in *Time:*

> Modern educators are proud of the fact that they have vastly improved the methods of teaching children to read. They estimate that today pupils can learn to read with more sense, more fun and in half the time that youngsters spent on the job 40 years ago. For this great improvement, research in child psychology is responsible (Oct. 23, 1939).

References

Adler, M. J. *How to read a book:* The Art of Getting a Liberal Education, New York, Simon & Schuster, 1940.

Efficient eye movements and a rapid rate of reading are no guarantee that one gets all possible benefit out of reading. Something over 99 per cent of college students, and most of their professors, could profit by following the advice of *How to read a book.*

Bean, K. L., *An Experimental Approach to the Reading of Music.* Psychol. Monogr., 1938, 50, No. 6.

Studies of the eye movements in reading music and methods of improving one's ability to read music are reported.

Buswell, G. T., *Fundamental Reading Habits:* A Study of Their Development. Supplementary Educational Monographs, 1922, No. 21.

Here may be found the detailed results of the study of eye movements summarized in this chapter.

Gates, A. I., *The Improvement of Reading.* New York, Macmillan, 1935.

Gates describes some of the techniques useful in teaching and in improving reading skill.

Gumbinger, R., *The Comparative Values of Three Methods in Improving Reading Achievement.* Unpublished M. A. thesis, University of Chicago, 1939.

Woodworth, R. S., *Experimental Psychology,* New York, Henry Holt, 1938.

Chapters XXIII and XXVIII contain authoritative and interesting summaries of the work on eye movements and reading.

CHAPTER 22

The Transfer of Training

HOW MUCH improvement in general memory is brought about by practice in memorizing poetry? How does studying mathematics influence one's reasoning ability? Does knowledge of Latin increase one's knowledge of English? In general, how widely does specific practice in one field transfer to related fields? This is the problem of transfer of training—a problem which has received almost every shade of answer from an unquestioning belief in the benefits of formal discipline to the practical denial that any transfer ever occurs. Experimental studies, giving answers in between these two extremes, have usually shown only very moderate amounts of transfer. Generally it has been reported that practice in one field has brought about some transfer to closely related types of activity, but neither the amount of transfer nor the range of activities over which the transfer was found have been very great.[1]

Results of transfer experiments. A few experimenters have found considerable transfer. Even more important, they have provided suggestions as to why unusually large amounts of transfer are sometimes obtained. Judd reported an experiment in which boys were taught to shoot at a target 12 inches under water. Some of the boys were given a full explanation of the refraction of light and how to correct for it in aiming. The other boys were not given this information. Both groups improved at about the same rate. When both groups had become proficient, the target was raised to a position only four inches below the surface of the water. The boys to whom refraction

[1] The literature is summarized in many textbooks of educational psychology. Since the topic has been a hotly debated one, some of the accounts are rather biased. Probably the best brief review is Chapter VIII of R. S. Woodworth's *Experimental Psychology*.

had been explained were far superior to the others, who had to learn all over again. Different methods of teaching had produced much greater transfer in one group than in the other.

In a very different sort of experiment Squire found that the habit of making neat arithmetic papers failed to transfer to other school subjects. In contrast, Ruediger found that if practice in neatness in one subject was accompanied by emphasis on the ideal of neatness, there was improvement in the papers written in other school subjects even though these subjects were never mentioned in the training. Again, a difference in amount of transfer followed a difference in the method of teaching.

Thinking over these results and the extent to which they differed from those of most studies, Woodrow asked:

> May not the general problem which is raised by these few citations be stated as the problem of the difference, with respect to the resulting transference, between unenlightened drill and intelligent teaching?

Woodrow set up the following experiment to test this hypothesis.

Three groups of subjects were given six memory tests. During the next four weeks one group was given no special practice or training in memorizing; this was the *control group*. A second group was given eight periods of drill in memorizing poetry and nonsense syllables; this was the *practice group*. The third group was given eight equal periods devoted to discussion of the principles of learning and techniques of memorizing, with some time spent in practicing the application of these principles to the learning of poetry and nonsense syllables; this was the *training group*. At the conclusion of the experiment all three groups were given parallel forms of the tests used at the start. By comparing the amount of improvement shown by the three groups it was possible to determine the relative effectiveness of the practice and training procedures.

The control, practice, and training groups included, respectively, 106, 34, and 42 subjects. The subjects were university

sophomores. The tests given at the beginning and end of the experiment were similar in form, but contained different material. They covered six types of learning material:

1. *Rote poetry.* Selections from *Mistress Gilpin* and *Alice Brand* constituted the initial and end tests.
2. *Rote prose.* Selections from Benjamin Franklin's *Autobiography* were used.
3. *Facts.* Each test consisted of 20 miscellaneous facts selected from a dictionary of facts.
4. *Turkish-English vocabulary.* Thirty words were included in each test.
5. *Historical dates.* Twenty little known dates were used in each.
6. *Memory span.* The number of consonants which could be correctly reproduced after they had been heard once was the score.

During the four weeks between the beginning and end tests the practice group was given 90 minutes of practice in memorizing poetry and 87 minutes of practice in memorizing lists of nonsense syllables. No directions were given as to the best methods of learning; the subjects were simply told to memorize the material given them. The schedule followed by this group is given in detail in Table 25.

TABLE 25

PROGRAM OF PRACTICE PERIODS FOR PRACTICE AND TRAINING GROUPS

Period	*Practice group*	*Training group*
I	20 minutes memorizing poetry	7 minutes listening to exposition of rules 13 minutes memorizing poetry
II	25 minutes memorizing poetry	7 minutes listening to exposition of rules 18 minutes memorizing poetry
III	28 minutes memorizing nonsense syllables	28 minutes listening to exposition and illustration of rules
IV	20 minutes memorizing nonsense syllables	5 minutes listening to review of previous period 15 minutes memorizing nonsense syllables

Period		*Practice group*		*Training group*
V	19	minutes memorizing nonsense syllables	9	minutes attending to "Blackboard talk" on meaning of secondary associations
			10	minutes memorizing nonsense syllables
VI	25	minutes memorizing poetry	25	minutes memorizing poetry
VII	20	minutes memorizing poetry	20	minutes memorizing poetry
VIII	20	minutes memorizing nonsense syllables	20	minutes listening to review of methods, and the situations in which to use them
Total	177	minutes	177	minutes
	90	on poetry	76	on poetry
	87	on nonsense syllables	25	on nonsense syllables
			76	on rules

Table 25 also outlines the training procedure followed by the third group. This group spent 76 minutes in memorizing poetry and 25 minutes in memorizing lists of nonsense syllables—a total of 101 minutes of practice as compared with 177 minutes for the practice group. The remaining 76 minutes for the training group was devoted to systematic study of various rules which previous experiments had shown to be valuable aids to learning. These rules were:

1. "Learning by wholes." In memorizing passages which are not too long, it has been found more economical to learn the passage as a whole than to break it up into smaller units, such as verses.

2. "Use of active self-testing." Spending a large part of the time in actively trying to recall the material, and then checking up on errors, has been found greatly superior to the more passive type of study in which the material is simply read and reread without trying to recall it until the study time is over.

3. "Use of rhythm and grouping." In the memory span test the consonants were read at a uniform speed. More can be remembered if they are grouped. For example, 573 492 837 is easier to remember than 5 7 3 4 9 2 8 3 7.

4. "Attention to meaning and the advantage of picturing, or—depending upon the individual—otherwise symbolizing the mean-

ing." The more associations one can think of, the more one can make the new material tie in with the things he already knows, the more vividly he can imagine it, the more easily will the learning proceed.

5. "Mental alertness and concentration." Again, an active attempt to learn gets better results than a more passive one.

6. "Confidence in ability to memorize." Confidence avoids the distracting effects of worrying about ability and about the coming tests.

7. "Use of secondary associations." In relatively meaningless material, such as nonsense syllables, learning is sometimes aided by using any associations, no matter how fanciful, which the material elicits.

The results of the study showed a clear superiority of the training group over both the practice and the control groups. The percentage by which the average performance of each group improved or declined during the experiment is shown in Table 26.

TABLE 26

PERCENTAGE GAIN OR LOSS OF THE THREE GROUPS ON SIX MEMORY TESTS

Test	*Control group*	*Practice group*	*Training group*
Rote poetry	—32.8	—29.1	—10.6
Rote prose	28.7	25.5	50.7
Facts (substance memory)	—4.9	—4.7	12.8
Historical dates	29.0	37.5	87.7
Turkish-English vocabulary	—0.6	3.4	55.2
Memory span	6.7	—5.7	20.3

Both practice and control groups improved on three tests and got worse on three. The lower scores at the end do not mean, however, that these two groups had actually become poorer learners; the tests of rote memory for poetry and memory for facts given at the end were more difficult than those given at the beginning.

By comparing the practice group gains with the control group gains a measure of the effectiveness of practice was obtained. The practice group gained a little more, or lost a little

less, than the control group on every test but one. The only one of the six differences which was statistically significant,[2] however, was on the memory span test in which the practice group did worse than the control group. It must be concluded that unguided practice did not significantly improve the ability to memorize.

The training group, in contrast, markedly surpassed both control and practice groups on every test, and all of the differences, shown in Table 27, were statistically significant.

TABLE 27

SIGNIFICANCE OF DIFFERENCES BETWEEN TRAINING GROUP AND CONTROL AND PRACTICE GROUPS

	Difference between means divided by the probable error of the difference between:	
Test	*Training and control groups*	*Training and practice groups*
Rote poetry	6.1	4.7
Rote prose	7.5	7.3
Facts (substance memory)	7.2	5.3
Historical dates	8.8	6.0
Turkish-English vocabulary	10.9	8.8
Memory span	9.0	17.0

It is to be concluded that the program followed by the training group made them significantly more efficient at memorizing.

In interpreting this improvement, attention must be given to the difference between the test and the practice materials. Practice was given on poetry and nonsense syllables, and on nothing else. Yet the trained group improved in their ability to learn

[2] In comparing two groups such as these it is necessary to ask whether the differences which are found might be due to chance variation. The most common way of answering this question is to divide the obtained difference, for example the difference between −32.8 and −29.1, or 3.7 (in the first row of Table 26), by the probable error of that difference. The probable error of the difference can be computed from the two distributions of scores. If the difference divided by its probable error is less than 4.5, it may be due to chance. If the difference is greater than 4.5 it is said to be statistically significant since chance variation is a most unlikely explanation.

The ratio of difference over probable error of difference will be used again in Table 27.

not only poetry, which was practiced, but five other kinds of material which were not. The difference between the improvement shown by the training and control groups on these other materials was in some cases even greater than it was on the practiced material—poetry. Most of the improvement on the six tests can be credited to the transfer or generalization of the improved learning methods which had been taught.

Further evidence of the generality of the transfer was found by correlating the scores on the different tests with each other. Had they all been measuring the same kind of memory, the correlations would have all been close to 1.00. Actually, the average correlation was only .38 which means that the six tests measured quite different aspects of memory. Thus the training given on poetry and nonsense syllables had transferred to five other kinds of material.

Woodrow concluded:

> The experiment shows that in a case where one kind of training—undirected drill—produces amounts of transference which are sometimes positive and sometimes negative, but always small, another kind of training with the same drill material may result in a transference, the effects of which are uniformly large and positive (171).

The Cox experiment. A quite different experiment which had a similar purpose and a similar result was conducted a few years later by J. W. Cox in England. Cox gave a number of subjects—school boys and adults—practice in one kind of mechanical assembly work and then tested their ability in other fairly closely related assembly jobs. The materials which Cox used are pictured in Fig. 41. When properly put together, these objects make an electric light socket. The tasks which Cox used consisted of assembling parts of the socket. One task required assembling parts *A* to *F*, in the figure, to make the container of the socket. Another consisted of the porcelain assembly—parts *G* to *M*. A third presented the subject with the porcelain already assembled, parts *A* to *F* unassembled, and some wire. The task was to make a completely assembled and properly wired socket.

Some subjects were given practice by having them assemble 440 socket containers; others were not given this practice. The practiced subjects became more skillful in assembling containers, but when both groups were tested on the remaining tasks

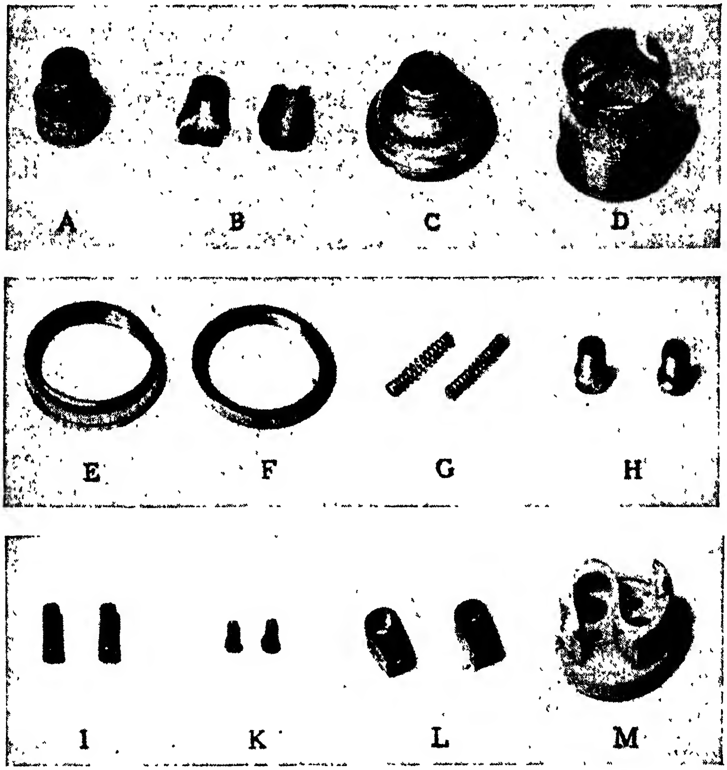


FIG. 41. Parts of the electric lampholder used in the "mechanical" and the "routine" assembling operation.

the results were such that Cox wrote: "Nowhere in our data is there any definite evidence of practice at one operation bringing about improvement at another" (146).

Cox, like Woodrow, distinguished between practice and training, and set up a course of training for one group of subjects. "The aim of the course was to impart a knowledge of certain general principles underlying the skillful handling of

assembly material and to provide specific exercise in applying them to *one* assembling operation, viz., the container operation" (166). This training was given in 11 lessons. Both discussion and practice were centered around five exercises:

1. Layout on bench, order of assembly, manner of holding parts, and so on.
2. "Eye observation exercises." What to look for.
3. "Finger observation exercises." What to feel for.
4. How to control attention and effort in the assembly work.
5. Practice in applying these principles to the assembly of containers.

Each lesson opened with a review of the points which had been made before. Attention was then given to the new point which was explained and demonstrated. The men were then given practice with special attention to becoming familiar with the new point. There was no emphasis on speed during the training. The last two or three days were devoted to a review of the chief points and to correcting individual errors. During the training course each man assembled 85 containers.

When this training group and the group which had practiced by assembling 440 containers were both switched to new and unpracticed tasks, the training group was superior in all five new tasks and *continued to excel still more* as further work progressed. With one exception the differences were statistically significant. Cox concluded:

> Skill, developed by the mere repetition of one manual operation, confers little advantage in the performance of other operations that may be subsequently undertaken. Where, on the other hand, repetition is replaced by suitable instruction, the skill thus developed at no additional cost in time tends to transfer to other operations over a fairly wide range of manual activity. This transfer is manifested not only in superior ability, but also in a superior rate of progress. The advantage thus conferred by training was obtained, in the present experiments, without any loss of efficiency during the training period (176).

The Katona experiment. Katona has recently reported a study of the relative transfer value of different methods of teaching college students to solve simple reasoning problems. As his experimental material Katona devised a number of geometric puzzles which can be worked with matches. Three samples are shown in Fig. 42.

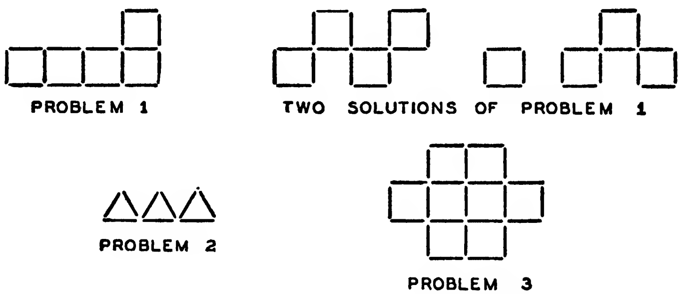


FIG. 42. Three of Katona's match problems.

In problem 1, three matches are to be moved so that there will be four instead of five squares. In an acceptable solution each match must be used as one side of a square, and each square must consist of only four matches. Two of the solutions are shown.

In problem 2, four triangles are to be made by moving three of the matches.

In problem 3, the eight squares are to be reduced to six by moving three matches.

Some subjects were shown the solution to one or two of the match problems. Others were told the general principles by which such problems could be solved. Still others worked the principles out for themselves through solving several simple problems which required different types of solutions.

Being shown the solution to one problem enabled the subjects to solve that problem, but did not aid appreciably in solving others. Being told the principles involved resulted in some positive transfer to new problems, but was not as effective as was working out the principles through examples or with the help of the experimenter. Once more, the amount of

transfer depends upon the method of teaching, being least when routine drill is used and greatest when the subject learns general principles and how to apply them in a variety of situations.

Results of above experiments. The studies by Woodrow, Cox, and Katona demonstrate unequivocally that proper training, through wide transfer, may result in improved ability in a whole range of tasks, while less effective training, or mere routine drill, may produce practically no transfer. The implications of these findings for education are obvious: to develop teaching methods which obtain the most transference is a practical and important challenge to all teachers.

References

Cox, J. W., *Manual Skill:* Its Organization and Development, Cambridge, Cambridge University Press, 1934.

Katona, G., *Organizing and Memorizing:* Studies in the Psychology of Learning and Teaching, New York, Columbia University Press, 1940.

Mursell, J. L., *The Psychology of Secondary-School Teaching.* Revised edition, New York, W. W. Norton, 1938.

Mursell gives a number of suggestions for increasing the transfer value of some of the school subjects.

Woodrow, H., "The Effect of Type of Training Upon Transference," *J. Educ. Psychol.*, 1927, *18*, 159-172.

The Chicago College Plan [1]

Who ever heard of college Freshmen and Sophomores asking for the more frequent administration of searching, penetrating examinations? Who ever heard of junior-college students requesting that the library be kept open longer hours and that the privileges of book withdrawals be enlarged? Who ever heard of junior-college students asking for extra discussion-section meetings in a course for which they receive no course credit? Who ever heard of junior-college students writing long papers that are not required and for which no credit or course-mark premium is offered? . . . Who ever heard of junior-college students daily discussing the most fundamental questions in philosophy, literature, biology, astronomy, economics, political science, and sociology at the dinner table and in the lounges of dormitories and clubs? . . . We have experienced each of these developments under the operation of our New Plan. (*The Chicago College Plan*, p. 106.)

YEARS ago when a student went to college he took the courses which the faculty had decided were good for him. He had little choice in his curriculum, and elective courses were at a minimum. As knowledge in various fields was broadened, new courses were introduced, usually as electives. Gradually the electives became so numerous that the original curriculum got lost in a maze of group requirements, alternative curricula, and free electives. Finally a point was reached in

[1] In this case, as in a number of others, the editor has asked the collaborator to describe an experiment conducted at his own institution. In the present study the experiment was on such a large scale that one might fail to recognize the carefully planned conditions under which the observations were made, certainly the essence of an experiment. Though the results are not primarily quantitative they do tell us a great deal about a very significant trial of methods for improving the effectiveness of college teaching and learning.

most American colleges, and still exists in many, where two students could get the same degree in the same year from the same college without ever having had any of the same courses.

Dissatisfaction with existing curriculum. A number of colleges and universities became so thoroughly dissatisfied with this chaotic condition that administrators and educational theorists agreed that drastic changes in the whole college organization were imperative. Among the changes that were suggested and made, the New Plan introduced at the University of Chicago in 1931 is one of the most thoroughgoing and successful. In 1935 Professor Boucher, Dean of the College, described the New Plan in the book *The Chicago College Plan.* In 1940 Professor Brumbaugh, who succeeded Professor Boucher as Dean of the College, brought the report up to date. The following account is abstracted from Dean Brumbaugh's revision of *The Chicago College Plan.*

The Chicago Plan. The main point of the Chicago plan is this: *"the major emphasis in the junior-college years should be devoted to breadth of general education"* (15). What constitutes a desirable general education could be, and at Chicago was, a hotly debated question. After much careful deliberation the program outlined for the junior-college years (the freshman and sophomore years) contained seven parts, five of which were required of all students, and two of which were electives.

Important features of Chicago plan. The most important feature of the plan is that these seven requirements are not stated in terms of courses which must be taken, but in terms of comprehensive examinations which must be passed. In order to secure the certificate showing satisfactory completion of the work of the College a student at Chicago must pass satisfactorily examinations, each six hours in length, covering seven fields of knowledge.

One of the five required examinations, the English qualifying examination, requires a demonstration that the student has developed acceptable and reliable habits of writing. The other four required examinations demand the attainment of the minimum essentials of factual information and an introduction to the methods of thought

> and work in each of four fields—the biological sciences, the humanities, the physical sciences, and the social sciences—such as may be expected of a student who has pursued through an academic year a general course at the junior-college level in each of the four fields. These five examinations represent a common core of educational experience and background for all students who complete the requirements of our College: they constitute the major part of our definition of the minimum essentials of a general education (19-20).

In addition, each student must pass two examinations chosen from the following list: Art, a laboratory course in Botany, Zoology, and Physiology, Chemistry, English, French, Geography, Geology, German, Greek, Italian, Latin, Mathematics, Music, Philosophy, Physics, a second course in the social sciences, or Spanish. There are three other requirements for the College Certificate: registration in the College for one year, and a minimum knowledge of mathematics and one foreign language. These latter two requirements are usually satisfied by work done in high school. In case a student lacks adequate training in mathematics or a foreign language, he must pass the appropriate examination in addition to the seven outlined above.

Syllabus provided for each examination. A syllabus, provided for each of these examinations, outlines the character of the work expected, lists required and suggested readings, and enables a student to prepare adequately for the examination whether he is in residence at Chicago or not. The examinations may be taken at any time they are given (usually twice a year) so that it is possible for a student to do his own reading, come to the university, and pass all seven examinations before he has even registered. Students capable of doing all of the work under their own guidance and capable of completing the College requirements in such a hurry are, of course, very rare. One of the advantages of the plan, however, is its flexibility in allowing each student to advance at his own pace.

The usual student spends two years attending courses corresponding to the seven examinations which he takes, some at the end of his first year, the rest at the end of his second

year. The most distinctive of these courses are those preparing students for the five required examinations. These are the introductory general courses in the biological sciences, the humanities, the physical sciences and the social sciences, and the English course. The four general courses are not all run on exactly the same plan, but they are similar enough so that the Introductory General Course in the Biological Sciences will serve as an adequate illustration. The official description of this course follows.

Introductory General Course in the Biological Sciences

The dominating objectives of the Introductory General Course in the Biological Sciences are: (1) to cultivate such skills and habits of scientific thinking as are exemplified by biology; (2) to describe and interpret the machinery of the organic world and the major concepts of biology; and (3) to provide such practical information as is desirable for a citizen in the modern world.

The content of the course is arranged in four main sections: (I) Variety and relationships among living organisms: a brief survey of plant and animal kingdoms, with emphasis upon man's probable ancestry. (II) The dynamics of living organisms: an analysis of how the living machine works, with particular emphasis upon the physiology and psychology of man in health and in disease. (III) Organic evolution, heredity, and eugenics. (IV) Ecology: the relation of living organisms to their environment and to each other; the problems of social organization in lower organisms.

The normal week's program comprises three lectures, one conference period, and approximately six hours of outside study. In addition, students are encouraged to visit the laboratory once each week. Here they are given opportunity to come in direct contact with the phenomena of biology through a new set of exhibits and demonstrations each week. The lectures are given by sixteen members of the regular staff of the division. Many of the lectures make large use of lantern slides and demonstrations. Students meet in small groups once a week for conference, and a conference leader is assigned to each group for the year. Conference periods are devoted to discussion, written work, and coaching in the technique of study.

Nature of introductory courses. The introductory general courses in biological sciences, humanities, physical sciences,

and social sciences are not introductory courses in the sense of preparing students for later specialization in the corresponding fields. Neither are they survey courses if that term is understood to mean a fairly rapid and fairly superficial study of the high points of a number of more or less closely related fields. Instead, each of the four courses was designed to give students who will *not* take further work in that field the information, habits of thought, and types of work considered important as part of the liberal education desirable in an enlightened citizen of the modern world.

Readings for each course provided. To achieve this aim a set of readings was selected for each course. In many cases appropriate books were not available, and new ones had to be written. Many methods have been used to help students understand and interpret their readings. Lecturers were chosen who combine facility in teaching with ability in some phase of the material studied. Talking motion pictures and demonstration laboratories were developed to illustrate the major phenomena of the biological and physical sciences. Trips to Chicago's Art Institute and museums are offered as means of illustrating the subject matter of the humanities course. Guided tours to industrial plants, depressed areas, and various social, economic, and political organizations provide similar first-hand observations for students in the social science course.

Discussion section leader. In each of the four courses each student is assigned to a discussion section leader. These men usually have the rank of Instructor or Assistant Professor, and are permanent members of the faculty. Each discussion section meets once a week. The major portion of the time is devoted to answering questions, clarifying misunderstood points, interpreting recent demonstrations, discussing the meaning of statements made by the lecturer or encountered in the readings, and otherwise helping the student to understand and to think critically about the material which he is studying.

Besides conducting his discussion sections, each leader is available a number of hours each week for individual conferences with his students. So much advantage is taken of these

conference opportunities that "there is no doubt that students have more opportunities for, and more actual, personal contact with, instructors under the New Plan than under the old one . . . there is provided for each student as much instructional assistance as he may need or desire" (59-60).

Students who have completed one of these general courses and who have passed the corresponding examination frequently go on to specialize in that field, but preparation for specialization is not the primary aim of the introductory courses. The chief purpose is to suit the needs of the non-specialist. Together, the four courses provide the bulk of the liberal education which Chicago considers desirable either as a basis for specialization or as the culmination of formal education for those who do not wish to become specialists. Regardless of what happens to him afterward, each student who completes the work of the College at the University of Chicago has this much of a liberal education: "an introduction to each of four large fields of thought, an essential minimum of proficiency in English usage, and a respectable minimum training in a foreign language and in mathematics" (25).

Chicago Plan an experiment in teaching. The experiments by Gumbinger, Woodrow, Cox, and Katona, reviewed in Chapters 21 and 22, all had the common purpose of testing the effectiveness of some method of teaching or learning. In each case the effectiveness was measured by comparing the results with those of a control group or a group trained by some other technique. The Chicago College Plan is also an experiment in teaching or learning, but on a much broader scale than the ones already reviewed. It might have been possible for Chicago to keep half of the 1931 Freshmen on the old free-elective system while the other half entered the newly-organized College. Had this been done, Chicago too would have had a control group by which to evaluate its new curriculum. But Chicago did not keep a control group, so success of its reorganized curriculum must be measured by other means. That might be done by comparing New Plan students with students whose work was done, at Chicago or elsewhere, under the older system.

This comparison was not made in a formal statistical fashion, although, of course, it was, and still is, made informally by the faculty. Brumbaugh quotes from eleven faculty reports, each of which expresses some reason for preferring the new plan to the old. The instructor in charge of one of the second-year laboratory courses in the biological sciences reported that students who have had the Introductory General Course in the Biological Sciences have "successfully done a larger amount of more difficult work than he ever dreamed could be expected from junior-college students" (126). During the first year of the New Plan the students in the humanities course were given an examination covering the work done up to that time. The examination was largely objective in character. Each instructor estimated in advance what the average score would be. It turned out to be higher than even the most optimistic instructor had guessed. But, how valuable and how representative are eleven faculty reports? If the entire faculty had been polled, how many would consider the new plan more successful than the old? Brumbaugh gives no figures. It may be assumed that a majority of the faculty prefers the new plan; otherwise it would probably not have been retained.

Brumbaugh writes: "The success of any plan of education must be ultimately determined by the students who are educated under that plan" (140). Changes in student work or student attitudes may both be appropriately used in this connection. Supporting the already cited faculty opinion regarding the quality of work are statistics on the reading done by New Plan students. For the Introductory General Course in the Biological Sciences the student rents a set of eleven books. A similar set for the Introductory General Course in the Physical Sciences includes seven volumes. The readings in the humanities and social science courses are taken from so many different sources that rental sets are not feasible. Instead, multiple copies are to be found in the College library. During the academic year 1937-1938 students in these two courses, on the average, withdrew two volumes per week from the library. The reading assignments have varied somewhat from year to year.

In 1934-1935 the requirements in the humanities course totaled 4224 pages; in the social science course 7894 pages were required. "The social science courses include two distinct types of works: some that are to be read intensively, as one expects to read a text, and others that are to be read more cursorily, as one reads a newspaper, merely for the sake of general impressions" (201).

Amount of student reading. The amount of reading done by freshmen and sophomores at Chicago should be compared with the amount done by junior-college students at other universities. No statistical comparisons are offered, but most college students will recognize that their own reading assignments are not as extensive as those quoted above.

Considering all available lines of evidence, it seems definite that students at Chicago maintain a higher level of intellectual work than do most junior-college students. The New Plan itself may be directly responsible for part of this difference, but in part it seems to be due to the fact that the New Plan attracts a superior group of students. Figure 43 shows a year-by-year comparison of the average Chicago freshman with the average freshman entering some 400 colleges and universities scattered over the United States. The curve shows the gradually increasing superiority of Chicago students in terms of intelligence test scores. The high caliber of student work at Chicago is undoubtedly due in part to the high quality of students attracted by the Chicago College Plan.

Student attitudes toward New Plan. Attempts have been made to measure the attitude of New Plan students toward the education they have received. One measure is perfectly objective.

> The percentage of Freshmen who have returned year by year as Sophomores is larger than the percentage of returning Sophomores prior to the introduction of the New Plan. Since we know that because of a prolonged depression the number of Freshmen who were prevented by financial difficulties from returning the following year is larger than in previous years, it seems that more students are

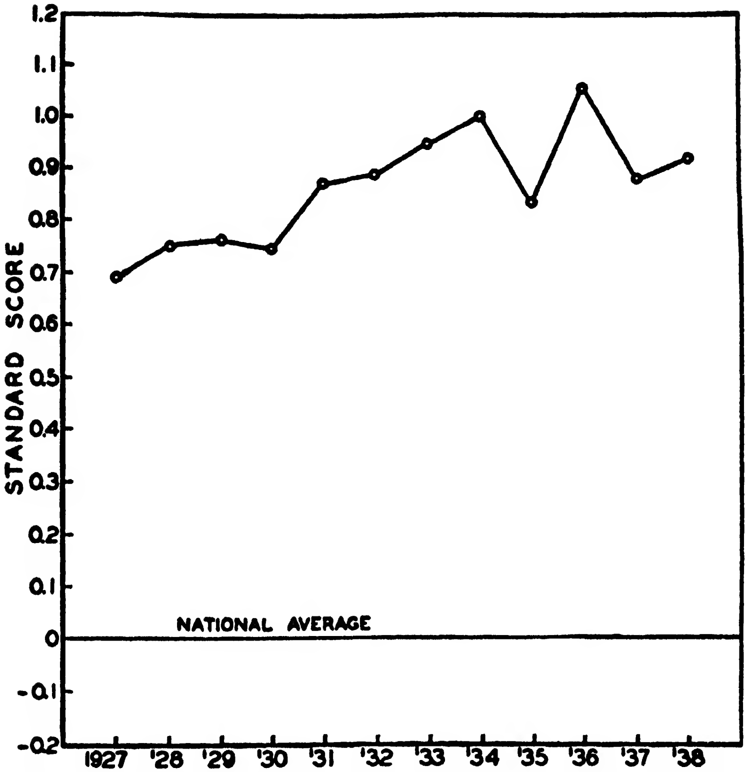


FIG. 43. Standard scores of median University of Chicago students in terms of national norms on the American Council on Education Psychological Examination.

In the four years before the beginning of the New Plan (1927-1930) the average entering freshman made a score of about .75. The average score rose steadily from the beginning of the New Plan until 1934. Since then it has fluctuated at a high level. A score of 1.0 is higher than that made by some 84 per cent of students in all of the colleges using the American Council on Education test. Another illustration of the meaning of the graph is to say that in 1934 84 per cent of the students entering *all* of the colleges involved made lower scores than the average student entering the University of Chicago.

satisfied with their education under the New Plan than under the old plan (133).

A second measure of attitude comes from both faculty and student reports of student-teacher relations under the new system.

Under the New Plan, with the award of credits, marks, and grade points taken away from the instructor, with the student free to attend class or absent himself at will, with courses offered solely with a view to assist the student in preparing himself for the demonstration of genuine educational attainment through the medium of comprehensive examinations, with the official Board examinations the only ones that count, and with these examinations scored by some one other than the instructor, it is evident that the traditional relationship between instructor and student is completely changed (109).

One example of this change is the absence of cheating in the class room. There is no point in copying another's quiz or term paper when neither counts toward a grade but both are solely opportunities for checking up on one's own progress.

Another example of this change is the absence of any need to cater to the individual instructor's whims. Since he has nothing to do with the grades, there is no reason for not adopting as critical an attitude toward the instructor as one can toward the text. Faculty reports show enthusiasm over the extent to which junior-college students can be critical, can think for themselves, and can show a high degree of evaluative judgment regarding the outside reading they do, the fairness of the quizzes and examinations they take, and the adequacy of the demonstrations, motion pictures, and lectures offered to them. The opportunity for criticism has a double value. It frees the student from any feeling of having to be careful not to antagonize the instructor, and it forces the instructor to do a better job of teaching. Both of these values have been realized.

A third attempt to measure student attitude was by means of a questionnaire sent to all students who had entered the College in 1931, 1932, 1933, 1934, and 1935, and who had completed the work for the College Certificate. Somewhat con-

densed, the chief results of one part of the questionnaire are shown in Table 28.

TABLE 28

PERCENTAGE OF 1065 STUDENTS GIVING THE INDICATED RESPONSES TO VARIOUS QUESTIONS IN THE EVALUATION QUESTIONNAIRE

Question	*Percentage answers in* *Biological Sciences*	*Humanities*	*Physical Sciences*	*Social Sciences*
1. Was the instructional material well organized?				
Yes	98	69	72	51
No	0.5	14	13	24
No answer	1.5	17	15	25
2. How was the quality of instruction in the lectures?				
Satisfactory	95	82	81	83
Unsatisfactory	3 1	15.5	14	15
No answer	1.9	2.5	5	2
3. How was the quality of instruction in the discussion sections?				
Satisfactory	81	80	70	56
Unsatisfactory	16	17	22	40
No answer	3	3	8	4
4. To what extent have you read books and articles in each field as the result of your experience in the course?				
Much	21	35	14	33
Some	47	36	35	47
None or practically none	31	28	48	18
No answer	1	1	3	2
5. Did the work help you to discover an intellectual interest?				
Yes	56	67	40	61
No	31	21	43	26
No answer	13	12	17	13
6. Did the work give you a greater satisfaction in living your life?				
Yes	83	86	66	77
No	8	6	18	12
No answer	9	8	16	11
7. Should every student be required to take the course?				
Yes	93	92	79	90
No	3	4	11	4
No answer	4	4	10	6
8. Did the comprehensive examination seem fair?				
Yes	89	71	80	76
No	5	18	10	11
No answer	6	11	10	13
9. How would you rate the course as a basis for understanding and interpreting problems of contemporary society?				
Of great value	54	62	31	87
Of little value	38	32	45	10
Of no value	5	4	19	2
No answer	3	2	5	1
10. Did the course teach you to be critical of slogan-thinking and careless use of the vocabulary?				
Yes	53	43	41	85
No	26	32	34	8
No answer	21	25	25	7
11. To what extent did the course help you to acquire habits of thought and methods of attacking problems that have been of value to you?				
Much	38	25	31	46
Some	34	38	35	29
Very little or none	26	34	29	22
No answer	2	3	5	3

One must always be cautious in interpreting the returns on a questionnaire. In this case 1924 blanks were mailed out and 1065 returned. Getting back 55 per cent of the blanks gives one a good deal of information, but still leaves one wondering how the 45 per cent who did not answer might differ from those who did. It is reasonable to suppose that those who hold their opinions most strongly are most likely to reply, and it is possible that those whose attitudes are favorable are more likely to reply than those with unfavorable attitudes. If these factors alone determined who did and who did not return the questionnaire, conclusions based on the returns would not fairly represent the entire group. But other factors also operate in determining which questionnaire blanks were filled out and returned. Some people were too busy to be bothered; others lost their blanks. Were these people more or less convinced of the success of the New Plan than were those who returned their blanks? No one knows the answer, but the best guess is that they probably did not differ greatly from the average of those whose opinions were recorded. Considering all of the selective factors, some of the figures in the table would undoubtedly be changed if replies had been received from everyone. The selection of those who replied from the whole list to whom qustionnaires were sent was, however, probably based on enough different variables so that few if any large changes in conclusions would be necessitated by receiving replies from everyone.

Replies to the fourth, fifth, and sixth questions indicate that the students in the main thought that they had received a fair amount of intellectual profit from each of the four general courses. Even the physical sciences course, the one with the lowest rating on these three questions, gave two-thirds of the students a greater satisfaction in their lives.

Replies to the last three questions indicate that the course in the social sciences was particularly successful in training students to adopt critical attitudes toward problems of contemporary society, in developing an understanding of those problems, and in developing useful habits of thought in dealing with

them. The course in the biological sciences was next most highly rated on these questions. The less obvious connection between the humanities or the physical sciences and contemporary social problems was reflected in the lower ratings of these two courses.

Probably the most significant question from the standpoint of evaluating the program as a whole is the seventh in the table —Should every student be required to take the course? At least 90 per cent of the students thought that everyone should be required to take the courses in the biological sciences, the humanities, and the social sciences. Only 79 per cent thought the physical sciences course should be required of everyone. Once more it would be interesting to have comparable information from other colleges. Lacking it, one can only guess. It seems probable that in few colleges would 90 per cent of the students agree on three courses which should be required of everyone.

Considering the questionnaire as a whole, there is no doubt as to the favorable attitude of most students. This evidence, together with the objective fact that more students return for the sophomore year, adds support to the conclusion, based on faculty opinion and examination results, that the New Plan is providing a more useful, more satisfying, and more thorough general education than the older system of free electives ever did.

Reference

Boucher, C. S., *The Chicago College Plan* (Revised by A. J. Brumbaugh), Chicago, University of Chicago Press, 1940.

Part Seven

VOCATIONAL GUIDANCE

By

E. G. WILLIAMSON

University of Minnesota

CHAPTER 24

Occupational Ability Profiles

AN UNDERSTANDING of the qualifications required for a particular type of work is basic to the selection of employees in industry, as well as in the selection of an occupation by a student. In the past this type of occupational information frequently was unverified and of questionable validity. An employment manager, for example, often relied merely upon general observation for information about the kind of training and experience necessary for successful work of a particular type. Likewise, students who faced the problem of choosing an occupation sought to learn from books, pamphlets and from persons on the job an understanding of the qualifications required for a particular type of work.

Limitations of subjective impressions. That this type of occupational information has serious limitations for both employment selection and vocational guidance is evident when one examines the basic assumptions underlying these cursory methods of collecting such data. It is apparent, for example, that while an engineer with years of experience in some branch of engineering may understand the basic operations of that type of work, he may not understand or be able to infer accurately the type of aptitudes required of those to be selected for such engineering work. Hazy impressions based upon one's own experience oftentimes prove to be fallacious when subjected to critical experimental investigation. What may be factors in the success of one engineer may not be factors in the success of another engineer. General information based upon subjective impressions is not sufficiently precise and reliable either for employment selection or vocational guidance. To continue our example, it is self-evident that an engineer must have facility and skill in the use of mathematics, but general observation

and even years of experience do not always tell us *how much* mathematical ability is required to become a successful engineer, or tell us the relative importance of this factor.

We may further exemplify this fallacy in another field. It has been said for decades that to be a successful lawyer one must be able to reason well. While no one would deny that lawyers must be able to reason, yet common observation does not tell us how much of this reasoning ability a successful lawyer must possess. To avoid this particular fallacy in methods of discovering the qualifications necessary to success in different occupations, industrial psychologists are developing a more precise and refined method which makes use of occupational ability profiles. Before citing an example of this method we may gain a broader understanding of the problem from an examination of some related experiments.

Intelligence and occupational level. For more than twenty years we have known that men employed in different occupations may be classified, according to the level of general intelligence, to form what has been called an occupational intelligence hierarchy. Lawyers, doctors and other men in the professions possess, *for the most part,* higher levels of intelligence than do men engaged in bricklaying, carpentry, and other skilled occupations. Figure 44 exemplifies this grouping and shows in schematic fashion for each group the relative proportions of the adult population classified in terms of general intelligence and the corresponding occupational level.

The reader should not conclude from a reading of Fig. 44 that *all* persons engaged in the professions are superior in general intelligence to those engaged in occupations, such as nursing, in the next lower occupational level. While the *average* level is higher for the one than for the other, there is "overlapping" in the distributions of tested intelligence. In part such overlapping is caused by errors in the measuring instrument. But a more important cause is found in the fact that general intelligence is only one factor, though a very important one, involved in occupational success. For example, a lawyer with high intelligence may be deficient in personality or motiva-

tion and thus fail to perform at the high level of which he is capable. Conversely an individual of average intelligence,

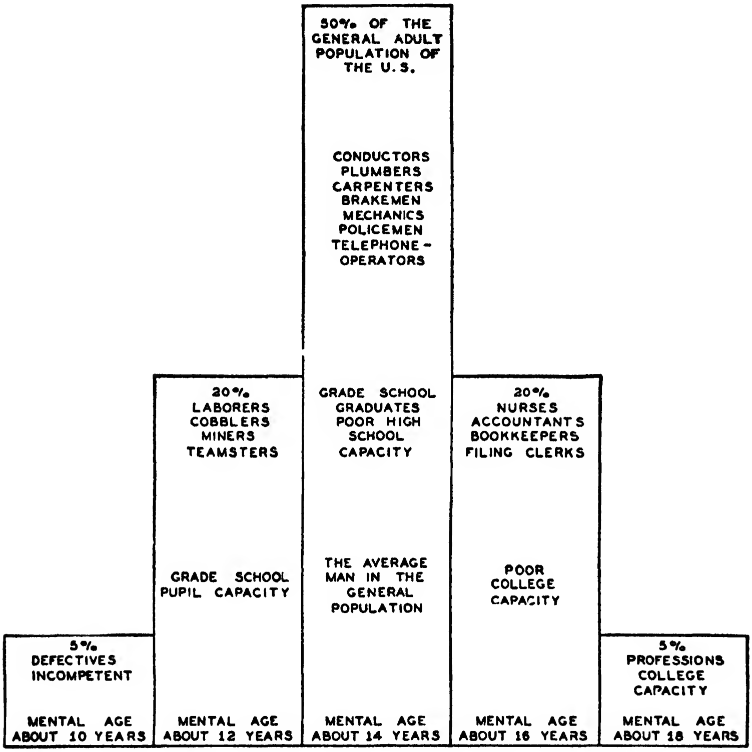


FIG. 44. Schematic graph, representing the approximate relative frequency in the population at large, of the five significant degrees of mental capacity, and indicating the most favorable type of occupational and educational career. In order to use this chart, a standard intelligence test should be given to an individual and his mental age determined. IQ may not be used in place of MA. (Hollingsworth, H. L., *Mental Growth and Decline*, New York, D. Appleton and Company, 1927.)

through diligent work and effective personal relationships, may achieve a level of success above that of other individuals of a comparable level of ability but without such favorable personal qualities. In predicting occupational success on the basis of

general intelligence one must modify the general prediction based upon group probabilities in terms of the personal characteristics of a particular individual.

The above chart shows a classification of occupations merely in terms of one type of aptitude. But more than this one type of ability is required to be successful in any given occupation. For example, bricklayers, although as a group relatively low in general intelligence, must nevertheless be high in other types of aptitudes. Therefore, an adequate description of the qualifications necessary for a job would require us to classify occupations in terms of several different types of abilities.

Minnesota Occupational Rating Scales. In the absence of more precise methods of measuring all of the relevant types of aptitudes, various judgmental classifications of occupations have been developed. An example is given from the Minnesota Occupational Rating Scales.[1] By means of judgments of informed experts, various occupations were classified in terms of five types of abilities: abstract intelligence, mechanical ability, social intelligence, artistic ability and musical talent. Jobs were rated in terms of each of these five abilities on a scale ranging from one as the highest degree of ability to six as the lowest. The profession of accounting, for example, requiring four years of college training, was classified as follows:

Occupation	*Abstract intelligence*	*Mechanical ability*	*Social intelligence*	*Musical talent*	*Artistic ability*
Accountant, auditor, abstractor —private or public; 4 yrs. college . .	2	6	5	6	6

This classification is an improvement over the older type of job description because it indicates the requirements in terms of more than one type of aptitude, but it still has the fault of not being based upon measured characteristics of men actually engaged in occupations. More precise classifications of occupations in terms of the abilities of the workers actually en-

[1] Bingham, W. V., *Aptitudes and Aptitude Testing*, New York, Harper and Brothers, 1937, pp. 365-380.

gaged in them is needed. The development of occupational ability profiles by means of testing and measurement techniques is an attempt to meet this need. The basic assumption behind these profiles is that successful men engaged in a particular type of work will resemble each other more closely with regard to required aptitudes than these same men resemble successful workers in unrelated types of work. Such profiles tell us not only what *types* of abilities are required for success in a particular type of work but also the *amounts* of abilities required. If these profiles are constructed in terms of the abilities actually possessed by successful men, then it is apparent that a particular individual's profile can be compared with that for any occupation under consideration to determine the extent to which he resembles successful men in this field. Other things being favorable, such as training and motivation, the individual whose profile of abilities most closely resembles that of men successfully engaged in the particular occupation will prove to be the most satisfactory worker. A similar line of reasoning is following in counseling students in the selection of an occupation.

CONSTRUCTION OF PROFILES

Occupational ability profiles may be constructed by means of three experimental procedures:

1. Selection of a battery of tests which differentiate successful members of a given occupation from the general population of adult workers.

2. Differentiation of this same group of workers from workers in another occupation.

3. Differentiation of members of the same occupation from each other in terms of degrees of success; for example, selecting a battery of tests on which the most successful sales girls in a department store score higher than do the least successful sales girls.

The development of an occupational ability profile. One of the most complete experimental investigations of the method of constructing occupational profiles is that reported by

Dvorak.[2] The following discussion is based upon Dvorak's study and the tables and figures are quoted from her monograph. Various successfully employed adult workers were given a standard battery of tests, the names of which are given in the tables below. Many of the statistical tables are not quoted here but may be found in the original study.

The first procedure in the construction of an occupational ability profile is, as indicated above, to determine the extent to which the tests used differentiate members of a particular occupation from the general population of adult workers. In Dvorak's study a carefully selected population of adults constitutes what is called the standard sample. In this particular study an attempt has been made to construct an occupational ability profile for men office clerks. The occupation selected for use in step two, namely, for the differentiation of this particular occupation from an unrelated one, was that of garage mechanic. Table 29 gives statistical data for these two occupational groups and the standard sample on the various tests used in this experiment.

An anaysis of the data given in Table 29 indicates that 90.4 per cent of the clerks reach or exceed the median score of the standard sampling of men on the test of educational ability and 98.2 per cent reach or exceed the median score of the general population on the test of clerical aptitude. Members of this particular occupation, therefore, possess on the average these two types of abilities in excess of men in general. Similarly, garage mechanics exceed the median scores of the standard sample of adult men on tests of mechanical abilities.

In Fig. 45 the median scores on the various tests are connected with straight lines to make a profile. The median score was selected because it is the most typical or representative score for these groups. It is apparent from inspection of this table that men office clerks as a group are highly differentiated from garage mechanics. The clerks score highest on tests of

[2] Dvorak, Beatrice Jeanne, *Differential Occupational Ability Patterns*, University of Minnesota Stabilization Research Institute, University of Minnesota Press, Minneapolis, Vol. 3, No. 8, February, 1935.

TABLE 29

MEAN SCORES AND STANDARD DEVIATIONS ON THE VARIOUS TESTS OF MEN OFFICE CLERKS, GARAGE MECHANICS, AND THE STANDARD SAMPLE OF MEN [8]

Test	*Office clerks*	*Garage mechanics*	*Standard sample*
Educational ability			
Pressey Senior Verification			
Number	110	102	423
Mean score	65.9	44.5	46.9
Standard deviation	15.0	17.3	20.9
Minnesota Clerical			
Number Checking			
Number	114	101	491
Mean score	136.8	85.6	83.1
Standard deviation	26.0	23.1	29.2
Dexterity			
O'Connor Finger *			
Number	113	102	488
Mean score	255.0	278.4	280.2
Standard deviation †	34.8	39.2	51.0
O'Connor Tweezer *			
Number	109	102	498
Mean score	323.1	352.2	360.6
Standard deviation †	47.0	54.4	61.5
Minnesota Manual: placing *			
Number	66	102	382
Mean score	224.0	236.6	240.4
Standard deviation †	18.3	22.1	30.0
Mechanical ability			
Minnesota Mechanical Assembly: Box A			
Number	69	102	494
Mean score	70.3	74.8	63.1
Standard deviation	13.6	11.0	19.3
Minnesota Spatial Relations *			
Number	113	102	489
Mean score	1148.8	1014.0	1262.0
Standard deviation †	252.7	209.9	308.5

* Scored in seconds.

† These measures represent one-half of the range between one sigma below and one sigma above the mean. The sigma distance below the mean is not exactly equal to that above the mean, for the statistical constants were calculated in terms of reciprocals on all the tests that are scored in terms of the time required to accomplish a given amount of work.

[8] Dvorak, Beatrice Jeanne, *ibid.*, p. 313.

general educational ability, clerical aptitude and dexterity. Conversely, the average mechanic exceeds the average clerk on tests of mechanical ability. Only 10.8 per cent of the mechanics reach or exceed the median score of the clerks on the Educational Ability Test. Only 3.7 per cent of the mechanics reach the median score of the clerks on the Clerical Test. On the other hand, 70.2 per cent reach or exceed the median score of the men office clerks on the Mechanical Assembly Test, and

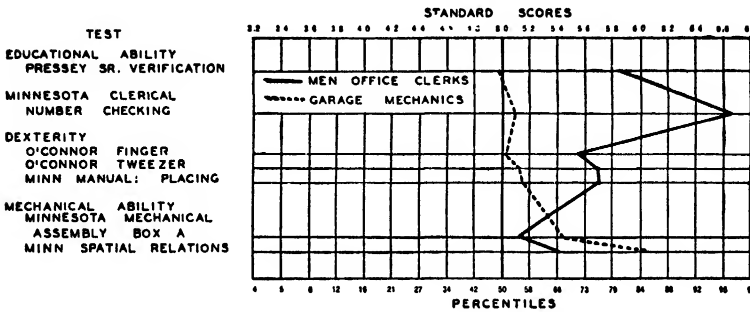


FIG. 45. Occupational ability patterns of men office clerks and garage mechanics (3, p. 314).

71.4 per cent reach or exceed the median clerk's score on the Spatial Relations Test.

A statistically more refined differentiation between the two occupational groups of mechanics and clerks was obtained by making use of mathematical weightings for the various tests in the battery. These weightings were determined by means of bi-serial correlation coefficients and indicate in a quantitative manner the extent to which each test differentiates the two occupational groups. The weighted scores on the various tests were added for each member of each occupation. These total weighted scores were then used to construct a more precise profile indicating the extent to which these two occupational groups differ on the tests used in this experiment. Figure 46 indicates in graphic form the extent to which garage mechanics and men office clerks possess different kinds of aptitudes.

An analysis of the data from which Fig. 46 was con-

structured, indicates that 96.0 per cent of the garage mechanics reach or exceed the median score of the men office clerks on the basis of the weighted total scores. This is in line with expectations since the weights were derived in terms of the extent to which each test was more characteristic of garage mechanics than of office clerks.

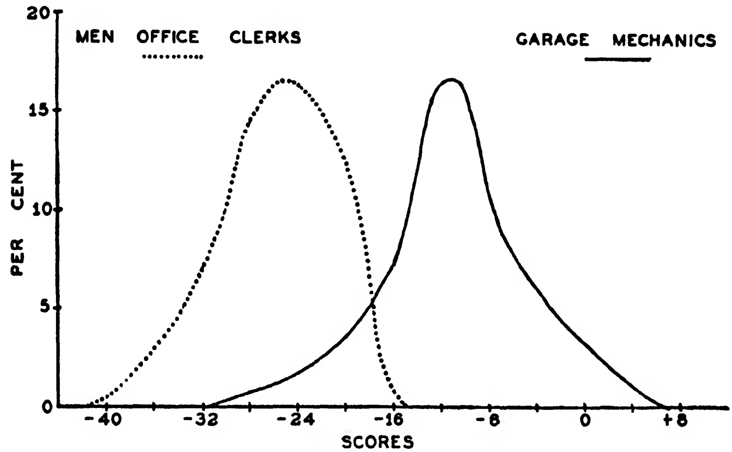


FIG. 46. Distribution of weighted composites of scores of garage mechanics and men office clerks (3, p. 317).

The third step in constructing an occupational ability profile, as indicated above, is the determination of the extent to which a test battery reveals differences among employed workers who differ in the degree of success in their particular type of work. Various studies are reported by Dvorak, one of which will serve to illustrate this procedure. Unfortunately no data are available on the garage mechanics and men office clerks used in the preceding steps; therefore, we shall turn to the analysis of a different occupational group. Figure 47 was constructed from the test scores of a group of women office clerks who were classified by their supervisors in terms of efficiency. These classifications range from group A, judged to be most efficient, to group E, judged to be least efficient. A comparison of the median test scores of these five groups indicates that the least

efficient workers, that is group E, are most clearly differentiated from the other four groups on all tests. Group D is less well differentiated but is still markedly different. The upper three levels of efficiency are not clearly differentiated by the tests used in this battery. In general, the most efficient clerks in this particular population are characterized by high levels of general educational ability and clerical aptitude; and poor

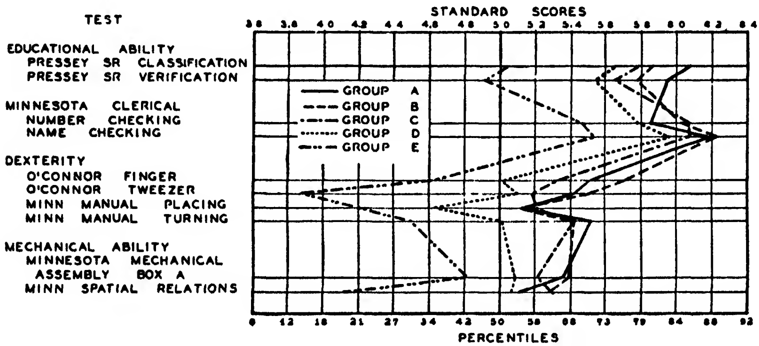


FIG. 47. Occupational ability patterns of different groups of women office clerks. (Dvorak, Beatrice Jeanne, *ibid.*, p. 324.)

performance on tests of manipulative abilities appears to be characteristic of the least efficient clerks. In general, then, the tests used to construct the occupational ability profile for women office clerks do differentiate employed workers according to level of efficiency. It should be noted that all of these clerks were employed and that presumably others who have extremely low clerical aptitude for this type of work had already been eliminated by the employers. Therefore, step three constitutes a very difficult test of the value of the test battery, in that it requires it to differentiate among successfully employed workers. Obviously this is a more difficult type of differentiation than one involving differences between successful and unsuccessful workers or differences between one type of occupational group and a totally unrelated type.

We have now discussed and illustrated the essential steps involved in the development of an occupational ability profile.

Certain supplementary procedures are involved but cannot be discussed at length in this chapter. Evidence from many sources indicates that occupational profiles are not universal in nature. That is, the standards maintained in one locality by a particular employer may differ from those maintained by another employer in a different locality. While one may, of course, set up uniform standards in constructing an occupational profile, yet the individual employer determines which standards are to be used in the selection and continued employment of his own workers. Dvorak presents evidence in her study which apparently indicates that within the locality in which she made her investigation these variations from employer to employer were not as marked as one might suppose. However, other studies, such as the one by Dodge, indicate the probability that there are marked differences from one locality to another.[4] Dodge also presents evidence that more valid occupational ability profiles may be constructed by means of minimum test scores required for successful employment in a given occupation as opposed to the median test scores, such as were used by Dvorak.

SUMMARY

The development of occupational ability profiles represents an attempt on the part of industrial psychologists to substitute objective standards for vague impressions in the selection of workers for employment in a given field. A demonstration of the validity of such a profile involves the determination of the extent to which aptitude tests differentiate successfully employed workers in a given field: (1) from adult workers in general, (2) from workers in an unrelated occupation, and (3) from each other in terms of level of efficiency as judged by supervisors and employers.

Not all of the qualifications required for success in any particular type of work are included in any occupational ability profile. Some qualifications, including certain personality traits,

[4] Dodge, Arthur F., *Occupational Ability Patterns*, Teachers College Contributions to Education, No. 658, Bureau of Publications, Teachers College, Columbia University, New York City, 1935.

cannot be measured at the present time because of lack of valid tests of those traits. In the selection of employees and in vocational guidance, therefore, trained counselors must supplement these profiles with additional facts about an individual. The use of occupational ability profiles, however, does provide an objective basis for the selection of individuals with greater probabilities of success than is afforded by the use of subjective impressions alone.

CHAPTER 25

The Measurement of Vocational Interests

Because of pressure from family and friends, most adolescents are motivated to make an early and sometimes rigid choice of an occupation in which they hope to find opportunity for the expression of what they conceive to be their interests. Early attempts to aid the adolescent in his choice of vocation involved two steps: the determination by the student of his own vocational interests through subjective analysis; and the selection of an occupation or a field of occupations which it was hoped would utilize these interests. A radical change in these procedures occurred when the psychologist brought to the guidance movement methods for the objective analysis of aptitudes and interests. Both experience and experiments with adults had made the psychologist skeptical of the accuracy and stability of an individual's estimate of his own aptitudes and interests. This skepticism led to attempts to develop methods of analysis which would provide dependable bases for vocational guidance.

The diagnosis of aptitudes. Much effort has been devoted during the past four decades to the perfecting of instruments for the diagnosis of aptitudes. Other experiments, dealing with the identification of vocational interests, have been performed. Many of these latter experiments were based upon the unproved assumption that the adolescent could diagnose his own interests in a valid manner and that the experimenter could take stated vocational choices at their face value.[1] While one may subjectively judge the intensity and direction of what he conceives to be his interests and occupational preferences, limited experience and information may often cause the choice of

[1] See Douglas Fryer's comprehensive review of these studies in *The Measurement of Interests*, Henry Holt, New York, 1931.

an inappropriate type of work for the expression of those interests. An occupation may demand the possession or later development of types of interests which one may not be able to identify and understand without years of experience. For the purposes of proper guidance, interests unrecognized and unknown to the adolescent must be identified before he has achieved the maturity and experience of an adult. The average life span is too short to permit one to devote many years to a tryout sampling of several occupations before making a choice.

We now understand that many adolescents are too frequently, and often unknowingly, influenced in what they conceive to be their work interests by such extraneous factors as pressure from parents and classmates, by vague desires to "do good in the world," to achieve financial success and similar goals which are not always closely related to their occupational qualifications. Fryer has shown that subjective estimates of the intensity and direction of interests of adolescents change so markedly from year to year that they provide undependable bases for vocational and educational guidance. More refined measurement techniques are needed to analyze the basic drives of the individual, many of which are not clearly recognized by him during the adolescent period.

EXPERIMENTS IN OBJECTIVE ANALYSIS OF INTERESTS [2]

Attempts at objective analysis of vocational interests began at the Carnegie Institute of Technology in 1919. Without tracing the detailed history of this interesting phase of psychology we may turn our attention to the fruition of this movement in research by Strong, who has perfected an interest inventory composed of 420 items [3] covering occupations, amusements, school subjects, activities, backgrounds of people and estimates of present abilities and characteristics. In filling out this in-

[2] This summary is confined to the inventory developed for men. A separate blank for women has not yet been studied in such detail.

[3] Strong, E. K., Jr., *Vocational Interest Blank for Men* (rev.), Form M, Stanford University Press, Palo Alto, 1938. The revised form consists of 400 items.

ventory the individual indicates whether he likes, is indifferent to, or dislikes each item in the test.

The test blank itself has gone through a number of forms. The current one, M, is divided into eight parts. Part I, dealing with 100 occupations, asks the subject to state whether he likes, is indifferent to, or dislikes each of the occupations listed. Part 2 deals similarly with 36 school subjects, Part 3 with 49 amusements, Part 4 with 48 activities, e.g., repairing a clock, and Part 5 with 47 peculiarities of people, e.g., optimists, sick people. Part 6 begins with a list of 10 activities. The three the subject most enjoys are to be given a *1* rating and the three he least enjoys a rating of *3*. This list is followed by another which includes ten factors affecting work, the names of 10 prominent men, and 10 positions in a club or society. This last list is to be checked in a manner similar to that employed with the first. Part 7 presents 40 heterogeneous pairs of items. The subject must state which pair member, if either, he prefers. Part 8 lists brief descriptions of 40 present abilities and characteristics. These are to be checked whenever they accurately describe the individual marking the test.

Constructing the test. If research had stopped at this point, the result would have been little more than a questionnaire of unknown meaning and value. Investigators who have stopped at this point have produced inventories which yield information as to what the individual thinks he is interested in, but they do not reveal the extent to which his interests coincide with those of criterion groups of adults. Building upon the work of previous experimenters, however, Strong standardized and validated his inventory upon groups of successful adults engaged for at least three years in various types of occupations and averaging 40 years of age. Except in a few groups the number of adults was at least 250. He determined for each of the three possible answers to each of 420 questions the statistical weighting to be attached to those answers.

The subject takes the test but once, yet each item may be scored or weighted differently for each occupation under consideration. Thus, if a subject states that he would like to be a

sales manager he receives a credit of +4 toward his life insurance salesman score, −2 toward his musician score, and a variety of weights ranging from +4 to −4 toward his other vocational scores. Similarly, a liking of public speaking when in school yields a weight of +2 toward the life insurance salesman score and a zero weight toward the musician score. In other words, by a fairly complicated statistical procedure it is determined how closely a subject's answers resemble those of successful men engaged in various occupations which appear attractive to the subject. Such a method of scoring permits an answer to the question, "To what extent do this individual's interests as expressed on this inventory agree with those of adults of average success in any particular occupation?" The basic assumption of such a statistical procedure is that other things being favorable (such as aptitude and opportunity) the individual will find his most satisfactory vocational adjustment in that type of work performed by men with similar patterns of likes and dislikes. By means of this inventory it becomes possible to determine not merely what an adolescent reports his interests to be but also the extent to which his likes and dislikes are similar to those of lawyers, accountants, etc.

The sum of the weighted answers to each of the 420 questions constitutes a total score which is then transmuted into the letter ratings A, B+, B, B−, C+, or C. The range of scores included in the A and B ratings differentiates the upper three-fourths of the criterion group of adults from the remainder of the group. A rating of C means that from 98 to 100 per cent of the criterion group of successful adults scored higher on the test.

At the present time the inventory can be scored for thirty-five different occupations.[4] As one might expect, these occupations are not all highly dissimilar with respect to the likes and dislikes of successful adults engaged in them. A grouping of 28 occupational groups into seven primary patterns, based upon an earlier study of Strong, is given below; this illustrates

[4] The revised inventory (1938) may also be scored for seven "group keys." See Table 30 below.

the clustering of occupations in terms of measured interests. The actual coefficients of inter-correlation which form the basis of this grouping may be found in Strong's published research.

TABLE 30

CLASSIFICATION OF OCCUPATIONS IN TERMS OF INTERESTS [5]

Group	*Occupation*
I	Chemist
I	Engineer
I	Farmer
I	Mathematician
I	Physicist
I	Architect
I	Dentist
I	Physician
I	Carpenter
I	Psychologist
I	Artist
IIA	Lawyer
IIA	Journalist
IIA	Advertiser
IIB	Life insurance salesman
IIB	Real estate salesman
IIIA	Minister
IIIA	Teacher
IIIA	Musician
IIIB	Personnel manager
IIIB	City school superintendent
IIIB	Y.M.C.A. secretary
IIIB	Y.M.C.A. physical director
IV	Accountant
IV	Office man
IV	Purchasing agent
IV	Vacuum cleaner salesman
V	Certified public accountant

The fundamental assumption underlying the construction of this test is that there is an appreciable similarity in the patterns

[5] Strong, E. K., Jr., "The Vocational Interest Test," *Occupations*, Vol. 12, April, 1934, p. 51.

of basic likes and dislikes of successful adults engaged in the same occupation. While there are undoubtedly idiosyncrasies which distinguish one lawyer, let us say, from another, yet basically lawyers will more nearly resemble each other with respect to reactions to their environment than they, as a group, will resemble engineers as a group. The following table derived by Strong illustrates this point: "The average engineer, for example, would dislike to be an advertiser, an actor, or an artist; is indifferent to being an army officer but would like to be an architect. The other five occupational groups have characteristically different patterns on these five occupational items." [6] The letter D indicated in Table 31 as the typical reaction of engineers to item "actor" is derived from a statistical weighting of the 420 questions in the inventory itself and not merely from the ten items appearing in the table below.

TABLE 31

TYPICAL LIKES (L), INDIFFERENCES (I), AND DISLIKES (D) OF SIX OCCUPATIONAL GROUPS [7]

Item	*Engineer*	*Lawyer*	*Life insurance salesman*	*Minister*	*Personnel manager*	*Accountant*
Actor	D	L	D	L	L	I
Advertiser	D	D	L	L	I	I
Architect	L	D	I	L	I	D
Army Officer	I	L	I	D	L	I
Artist	D	D	D	L	I	D
Algebra	L	D	D	I	D	L
Agriculture	I	D	L	L	I	I
Arithmetic	L	D	D	D	L	L
Art	I	D	I	I	I	D
Bible Study	I	L	L	L	I	D

Stability of interests. Psychologists are concerned with the question of the stability or permanence of the phenomena they are measuring. Consequently, they attempt to develop measuring instruments which will yield consistent results when used at different times with the same individuals. Essentially, this

[6] Strong, E. K., Jr., *ibid.*, p. 50.
[7] Strong, E. K., Jr., *ibid.*, p. 50.

is a question of the reliability of the measurement of any particular trait. Reference was made above to the fact that an adolescent's estimates of his own vocational interests fluctuate so markedly that one can have little confidence in these unverified expressions as the basis for planning an occupational career. If Strong's test had no greater degree of permanence or stability, then it would not be superior as a basis for vocational guidance to the technique of merely asking a student what he wants to become when he reaches maturity. An instrument for measuring basic interests is needed which is not colored greatly by the adolescent's changing estimate and interpretation of these interests. It may well be that the basic interests themselves do change; in which case our measuring instrument must reveal that fact and guidance based upon that instrument must be correspondingly modified. Table 32, con-

TABLE 32

PERMANENCE OF VOCATIONAL INTERESTS FOR ONE- AND FIVE-YEAR PERIODS—SAME INDIVIDUALS TESTED TWICE [8]

	Coefficients of correlation			
	One year		*Five years*	
	11th vs. 12th grade *	*Freshmen vs. sophomores in college*	*Seniors in 1927 vs. 1932*	*Same corrected for attenuation* †
Number of cases	100	247	223	223
Engineer scale	.828	.906	.844	.898
Physician scale	.726	.845	.778	.880
Lawyer scale	.652	.753	.769	.857
Minister scale	.678	.786	.643	.702
Average	.721	.822	.759	.834

* "Internal evidence suggests that a small minority did not take the task of filling out the blank seriously or else that there is surprisingly high permanence of interests at this age for most boys and very little permanency for the remainder."

† "The average reliability of 27 occupational scales is .878 (odd-even technique)."

structed by Strong, indicates the extent to which his instrument yields consistent or reliable measurements of vocational interests.

[8] Strong, E. K., Jr., "The Vocational Interest Test," *Occupations*, Vol. 12, April, 1934, p. 54.

The average reliability coefficient (total of odd-numbered items versus total of even-numbered items) for the various scoring keys on this inventory is .877 corrected).[9] This is not as high a reliability as is found in the best tests of intelligence and in certain aptitude tests, but it is far more stable than subjective estimates or interpretations of vocational interests.

Another test of the stability of interests has been made by Strong using as subjects 223 students at Stanford University who filled out the interest blank during their senior year and again five years later. The average retest reliability coefficients after five years for twenty-one scoring keys was .84 (corrected). There is some indication that these reliability coefficients would not be as high for younger age students.

A more detailed analysis of the stability of these seniors' interests was made by comparing the letter ratings on the original test and the retest five years later. Strong summarizes as follows:

> . . . 63.3 per cent of the ratings will be identical each time and 84.6 per cent of the ratings will agree within one rating of each other.
>
> . . . an A rating will be a B rating or higher in ninety-eight per cent of cases and a C rating will be a B rating or lower in ninety-six per cent of cases.[10]

Additional evidence of the permanence of interest, this time in terms of changes which occur from age period to age period, is found in a study of interest maturity. In this experiment Strong compared the measured interests of 472 representative fifteen-year-old boys (from California) with those of 632 fifty-five-year-old men successfully engaged in professions or business.

The results indicate that the interests of men between the ages of twenty-five and fifty-five are essentially alike. There

[9] See p. 14 of the 1938 Manual for this test.

[10] Strong, E. K., Jr., "Permanence of Vocational Interests," *J. Educ. Psychol.,* Vol. 24, May, 1934, p. 343.

are, however, noticeable changes in interests from fifteen years to twenty years. Some of these variations in interests between the younger and the older age groups are very marked, as illustrated by the responses to the following item (No. 187 in the inventory):

Public speaking	*Per cent who like the item*[11]
Age 15	20
Age 55	45
Difference	25

A similar analysis was made for every question in the inventory and then a scoring weight was given to each question which differentiated these age groups. The result was an interest maturity score which indicates the extent to which the individual filling out this inventory resembles the average adult in his general patterns of likes and dislikes. By means of this new scoring key a counselor may judge whether a student's interests are "adolescent" (unfocused) or crystallized. This maturity score had an odd-even item reliability of .94 for 100 fifteen-year-old subjects; .92 for 100 subjects from populations 25, 35, 45, and 55 years of age; and .97 for a similar population of 15- to 55-year-old subjects.

Validity of life insurance scale. Some years ago, Strong[12] studied the relationship between scores on the life insurance salesman scale and the annual amount of paid-for business sold by a number of salesmen. The possible test ratings ranged from A through C, A including the interests of 75 per cent of successful salesmen; B+, 16.1 per cent; B, 6.7 per cent; and C, 0.9 per cent. Table 33 presents the annual paid-for sales of Strong's subjects. The table should be read as follows: 85 per cent of agents with A ratings write insurance worth $100,000 or more annually; 51 per cent of those with B+ ratings write this amount, etc.

[11] Strong, E. K., Jr., "Interest Maturity," *The Personnel Journal,* Vol. 12, August, 1933, p. 81.

[12] Strong, E. K., Jr., "Interests and Sales Ability," *The Personnel Journal,* Vol. 13, December, 1934, pp. 204-216.

TABLE 33 [13]

RELATION OF INTEREST SCORE TO SUCCESS AS LIFE INSURANCE SALESMAN

Rating	*Per cent selling $100,000 or more annually*	*Per cent selling $150,000 or more annually*
A	85	67
B+	51	43
B	44	21
C	25	6

From these data it seems clear that vocational interest, as measured by the Strong Vocational Interest Blank, has a definite relation to vocational success in the important area of life insurance salesmanship. Similar relationships have not been established for men in other occupations.

Strong is now revising his maturity scale and has also published an *occupational level* score indicating the extent to which an individual's interest pattern resembles that of laboring men, on the one hand, and business and professional men earning at least $2500 a year, on the other.

THE ORIGIN OF VOCATIONAL INTERESTS

At the present time there is insufficient research to enable us to indicate the origin of interests or to tell how a student without a desired pattern may develop one. Some studies at the early and late adolescent periods indicate the relative stability of patterns of interests once they have been developed or crystallized. The environmental factors producing these patterns of interests are, however, unknown. From a common-sense point of view, an individual's work experiences would be acceptable as a determining factor, but there is much evidence to contradict this hypothesis. If an individual must actually do the work of a lawyer before acquiring the interests of a lawyer, then it would be impossible for an adolescent relatively un-

[13] Table IV is constructed from data given in non-tabular form in Strong's article. See above footnote No. 12. For further data of a similar type see Bills, M. A., and Ward, L. W., "Testing Salesmen of Casualty Insurance," *The Personnel Journal*, Vol. 15, June, 1936, pp. 55-58.

familiar with the detailed procedures in the occupation of law to receive an A rating on the Strong interest test. As a matter of fact, a large number of young individuals do receive high ratings on various occupational keys before they have had more than general acquaintance with the occupations. For example, Carter and Jones [14] discovered that this interest inventory apparently indicates broad trends for high school boys in the same way that it does for adults. A large number of these high school students received high ratings, A or B+, on various occupational keys; approximately two-thirds of them rated above C in their chosen occupation. Apparently, then, the interests measured by this inventory may precede actual experience in a particular occupation. An inspection of items in the inventory itself will indicate that these results should be expected, since these items are not restricted to interests in the detailed operations of a particular type of work but rather sample reactions to broad categories of experience. A second line of evidence indicating that actual occupational experience alone does not explain the origin of interest patterns as measured by this inventory is found in the fact that adults who have actually been engaged in a particular occupation for a number of years sometimes do not receive high ratings on the occupational keys most directly related to their type of work. This point was established by a study of the scores and letter ratings of Strong's criterion groups of successful adults. While it is evident that many of these adults, for example engineers who receive C ratings on the interest inventory, are maladjusted in their work, this merely reinforces the conclusion that years of a particular type of experience do not necessarily force one to acquire a pattern of interests held by one's associates.

SUMMARY

It has often been assumed that a vocational choice made by an adolescent is a direct and valid expression of his vocational

[14] Carter, H. D., and Jones, M. C., "Vocational Attitude Patterns in High School Students," *J. Educ. Psychol.*, Vol. 29, May, 1938, pp. 321-334.

interests. That this is not true in all cases is indicated by the fact that these choices change markedly from year to year. Objective methods of measuring the basic interests themselves without regard to the individual's interpretation of those interests have long been needed. It is quite probable that these basic interests are relatively permanent once they have become crystallized during the adolescent period. But the individual's identification of these interests and the choice of an occupational field which will utilize these interests is a difficult task. While the basic interests themselves may be relatively permanent, yet there may be instability in the identification of this pattern by the individual. Moreover, an individual may possess several independent patterns of interests, with the result that he shifts his attention from one pattern to another and thus appears to have no permanent pattern. The development of the Strong vocational interest inventory has made possible the identification and diagnosis of vocational interests. This inventory permits a determination of the extent to which an individual's pattern of basic likes and dislikes resembles those of successful adults engaged in any particular occupation. If an individual possesses other required qualifications, then his probability of successful vocational adjustment will be greatest in an occupation requiring likes and dislikes similar to those he possesses. The construction, standardization and validation of an inventory designed to yield an accurate analysis of basic likes and dislikes have been described in the preceding paragraphs.

CHAPTER 26

Predicting Scholastic Success in College

SINCE the time of the tremendous increase in the number of students enrolled in colleges following the first world war, admissions officers have faced the problem of sifting out from the large number of applicants those individuals who appear to have sufficient aptitude for successful work in college courses. Essentially the perfecting of means of identifying potentially successful students is a psychological problem. Attempts were made, therefore, to borrow from psychologists those testing techniques perfected prior to or during the first world war for the selection of employees in industry and for the classification of men in army occupations. The college admissions officers needed similar testing techniques, valid and objective in nature, to identify students who would probably fail, before admission to college. Unless such valid techniques could be used before students actually enrolled in college, no other selection method was available to reduce the number of students who failed in the first quarter or first semester.

A sound program of higher education. This need for new psychological techniques presented a real social opportunity for psychologists to contribute to the better development of a sound program of higher education. It seems reasonable that American colleges should use their resources for the training of those individuals who have the capacities to perform the types of work for which the colleges have been established. Just as an industrial plant should use its facilities to produce a product meeting a social need, American colleges have a public responsibility to graduate their students well trained in various specialties. Our society has given to colleges a public responsibility for the preparation of professional men of high competence. In

this sense the colleges are obligated to protect society from incompetent workers by graduating only well-trained ones.

The problem of selecting students with high probabilities of scholastic success is similar to that faced by industry and business in selecting employees with high probabilities for success. However, there is a difference in the problems faced by industry and the colleges. The college must be interested in the personal development of students as an important phase of the production of competent workers for society. Industry, on the other hand, is interested primarily in using employees to produce goods for sale to the public. The welfare of the employee is of concern to the employer chiefly insofar as it affects the production of high quality goods.

Psychological testing in colleges. The reader may be interested in more extensive reading in the history of psychological testing in colleges. At this point, however, we can only point out that psychologists' experiences during the World War were rapidly adapted to the attack upon the problem of selection of students, with the result that thousands of research studies have been made and published. To illustrate these many researches we have selected one or two from the many made by J. B. Johnston, a neurologist, who brought scientific research methods into the problem of college admissions with the able assistance of D. G. Paterson, a psychologist trained in industrial psychology and in testing in the army. We select Johnston's experiments because he worked steadily on the problem of developing selection methods from 1914 until his retirement in 1937. During this period a comprehensive attack upon the problem of "student mortality" or scholastic failures was made, involving not only the use of psychological tests but also the development of a counseling program, the instituting of certain reforms in teachers' grading, such as the greater use of objective tests and basic changes in college curricula.

ACCURACY OF PREDICTIONS

The research problem attacked by Johnston was how to identify by objective means those applicants for admission

whose previous records indicated high probabilities of successful work. As is true of other colleges, "success" in Johnston's college [1] was defined as maintaining an average grade of C or higher in class work. A subsidiary problem had to do with the success of students who transferred to professional schools, such as law or medicine. It was not anticipated that a standard technique or a standard level of ability would be applicable to all colleges. This principle of different standards of aptitude required for different colleges resembles closely the principle discussed in a previous chapter on occupational ability profiles with regard to differing standards for different communities and industrial establishments.

A psychological verbal test. A psychological test composed chiefly of questions sampling a student's familiarity with meanings of words was constructed by Paterson and used by Johnston. This type of test formerly was classified as a form of intelligence test but is now considered one of the tests of scholastic aptitude. In addition to the use of this type of test, Johnston discovered early that a college student's success could be predicted with a fair degree of accuracy by means of his average grade in high school transmuted into a percentile rank. A percentile rank of 50 was made equivalent to the average grade received by the average senior in a particular high school while in that school. By empirical rather than by elaborate statistical means Johnston found that a simple arithmetic averaging of a student's percentile rank on the psychological test with his percentile rank in high school grades yielded the most accurate prediction of college work. The result of this averaging of the two percentile ranks was called "college aptitude rating" (C.A.R.). In a later section of this chapter we shall see that many changing conditions in Johnston's college later caused him to vary the weighting given to these two variables.

College aptitude rating tests. Table 34 summarizes Johnston's research studies on freshmen enrolling in college from the years 1923 to 1927 inclusive. During that period the scho-

[1] The College of Science, Literature and the Arts, University of Minnesota.

lastic records of 2212 freshmen were studied. The table is interpreted as follows: A total of 88 freshmen enrolled in college with C.A.R.'s between 96 and 100. Of these 88 students, 85 or

TABLE 34

COLLEGE SUCCESS MEASURED BY COLLEGE APTITUDE RATING. FRESHMEN ENTERING 1923 TO 1927 INCLUSIVE [2]

Number and Per Cent Receiving Satisfactory Grades

C.A.R.	*Number*	*Number satisfactory*	*Per cent satisfactory*	
96-100	88	85	96.6	94.4
91- 95	125	116	92.8	
86- 90	156	121	76.3	71.9
81- 85	135	98	69.5	
76- 80	137	89	65.	
71- 75	123	69	56.1	42.
66- 70	131	57	43.5	
61- 65	144	66	45.2	
56- 60	144	57	39.6	
51- 55	150	45	30.	
46- 50	141	34	24.1	18.
41- 45	129	20	15.5	
36- 40	137	19	13.7	
31- 35	116	13	11.2	9.2
26- 30	90	6	6.4	
21- 25	76	1	1.3	1.1
16- 20	85	2	2.3	
11- 15	47	.	..	
6- 10	44	..	..	
1- 5	14		..	
	2212	898		

96.6 per cent received an average grade of at least C during the first year in college. The last column to the right of Table 34 indicates the percentage of successful students for various levels of C.A.R. An inspection of this column yields the clearest meaning of this table. If the reader will think of these percentages as indicating the probabilities of success of students of various levels of C.A.R., then it is clear that, in these par-

[2] Johnston, J. B., *Who Should Go to College?* University of Minnesota Press, Minneapolis, 1930, p. 21.

ticular years, an applicant for college should have a C.A.R. of approximately 70 before he has a fifty-fifty chance of maintaining an average grade of C or higher. If a student's C.A.R. is below 25, then he has only slightly more than one chance out of 100 of doing successful work in college. In fact, only three students with a C.A.R. below 25 did satisfactory work during this particular period of years. In general, the higher the level of C.A.R. the higher the probabilities of success. For students with C.A.R.'s between 36 and 50, the chances for success in college are about one in six. Of all the students who do satisfactory work in college, 89.4 per cent have C.A.R.'s above 50, whereas only 10.6 per cent have C.A.R.'s below 50. Of all the satisfactory students 99 per cent had C.A.R.'s above 30.

It is apparent that, in terms of this particular experiment, being an average student in high school does not yield a high probability of being an average student in college. In fact, one must be above average in high school work to be average in college work. This principle holds whether one talks about aptitude as measured by psychological test or whether one talks about aptitude as indicated by quality of work done in high school studies. A number of factors usually unknown to high school seniors provide the basis for this generalization. College class work is more difficult than that of high school because of the greater amount of material the student is expected to learn, the shorter time in which he is given to learn it, the greater responsibility he is given for self-directed study, and the more difficult examinations given by college teachers. The mere possession of aptitude is not sufficient to indicate high probabilities of success in college work. A more intense motivation to learn more difficult material is required of college students. Moreover, a college student must make a radical readjustment in his work habits almost immediately upon enrollment in college. If he does not make a rapid transition from high school to college, then it is likely that he will do such unsatisfactory work during the early part of his residence as to preclude the possibility of developing new habits of studying before he is asked to withdraw from college.

A crucial test of Johnston's early study, based upon the classroom work of students during the first year of residence in college, is found in a follow-up study made ten years later of a large proportion of the students included in Table 34. It was anticipated that some students who did unsatisfactory work during the first year might later make a satisfactory adjustment upon returning to college after a period of enforced absence. Table 35 gives the results of this follow-up study in terms of the percentage of students at each level of C.A.R. who ultimately graduated from the college in which they first enrolled. A comparison of Tables 34 and 35 reveals the fact that relatively few of the students who did unsatisfactory work during the first year in college were able to readjust themselves and graduate with a satisfactory standing. This is indicated by a comparison of students with C.A.R.'s below 25—1.1 per cent in Table 34 and 4.5 per cent in Table 35. The reader will note that at higher levels of C.A.R. the two percentages differ markedly. These differences are caused by the fact that a large proportion of the students at the high levels do not graduate from the Arts College but transfer at the end of two years to the professional colleges, such as law, medicine, education and dentistry. These transferred students are not to be classified as unsuccessful students. A further analysis not included in Table 35 indicates that a large percentage of these transferred students with high levels of C.A.R. succeeded in professional colleges and, on the other hand, students transferring with low levels of C.A.R. succeeded to a much smaller extent.

In Table 35 are included thirteen students with C.A.R.'s below 26 who transferred from the Arts College to professional schools and who ultimately graduated. One student graduated from Business, two from Dentistry, three from Dental Hygiene or Dental Nursing, four from Education, two from Engineering or Architecture, and one from Law. Only one of the thirteen students who eventually graduated from these professional colleges had an average grade of C in his Arts College subjects. Only five of the thirteen students graduated from a professional college which requires an average grade of C for gradua-

tion. All thirteen met the requirements for graduation of their respective professional colleges, but in courses taken in the Arts College, only one met the standard of a C average used in the

TABLE 35

PREDICTIVE VALUE OF C.A.R. WITH REGARD TO PERCENTAGE OF STUDENTS GRADUATING FROM THE ARTS COLLEGE [3]

College aptitude rating	*Per cent graduated*
96-100 91- 95	51.6 *
86- 90 81- 85 76- 80	35.1
71- 75 66- 70 61- 65 56- 60 51- 55	25.4
46- 50 41- 45 36- 40	19.4
31- 35 26- 30	13.9
21- 25 16- 20 11- 15 6- 10 1- 5	4.3

* The other 48.4 per cent includes those who canceled with C or higher average grades, or who graduated from another college or professional school.

above study. Some of these students with low ratings required longer than the normal period of college residence to complete their professional work. These results would seem to indicate that Johnston's C.A.R. has greatest validity for predicting Arts

[3] Johnston, J. B., and E. G. Williamson, "A Follow-up Study of Early Scholastic Predictions in the University of Minnesota," *School and Society,* Vol. 40, No. 1040, December 1, 1934, p. 733.

College grades and less for predicting success in professional schools which did not maintain comparable grading standards.

CHANGING CONDITIONS AFFECTING ACCURACY OF PREDICTION

Related research studies made more than a decade after Johnston's original study indicate that accuracy of prediction is not a constant phenomenon. Changes in grading standards, content of the courses, and many other factors produce variation in the predictive accuracy of any psychological test. For this reason a continuous program of research must be made if admissions procedures are to select students with high probabilities for successful work. Sometimes the factors producing these changes in predictive accuracy are unknown. Table 36 illustrates changes which may occur in the predictive accuracy of any psychological test and also illustrates the fact that the standard for success in college varies from period to period. The psychological test used in this later research was the same one used by Johnston in his early study. It will be seen in Table 36 that for approximately the same number of students,

TABLE 36

CHANGES IN MEAN SCORE, SIGMA, AND CORRELATION WITH SCHOLARSHIP OF THE 1926 FORM OF COLLEGE APTITUDE TEST[4]

	Total scores			
N	*Mean raw score*	*Mean percentile rank* *	*SD raw score*	*r*
792	248.0	53	58.3	.50
993	254.5	56	58.7	.45
951	278.9	72	48.9	.39
827	281.0	73	52.1	.43

* Percentile norms based on all Arts College freshmen enrolling in September, 1926.

the average score on the psychological test changed markedly from the period 1926 to 1935. This change is equivalent to an increase in average score from a percentile rank of 53 in 1926

[4] Williamson, E. G., "The Decreasing Accuracy of Scholastic Predictions," *J. Educ. Psychol.*, Vol. 28, January, 1937, p. 2.

to 73 in 1935 with a corresponding decrease in the standard deviation or spread of total scores around this mean.

Partly as a result of this decreasing spread of test scores the correlation between scores on this test and grades received by students in college also changed from .50 in 1926 to .43 in 1935. Correspondingly, other data not included in Table 36 indicate that the average high school percentile rank increased from 62 to 73 with a decrease in standard deviation from 29.4 to 22. Other statistical analyses not given here indicate that probably the change in test scores and high school rank are not the sole causes for the decrease in the size of the coefficient of correlation. It is suggested that changes in grading standards and content of courses as well as the development of a more efficient personnel program may have contributed to the changes in accuracy of predicting students' grades.

Predicting success in college work. A more recent prediction study for students enrolled in this same college indicates that the use of another psychological test in combination with high school grades yields almost as accurate prediction as that obtained by Johnston in his early studies. However, a new table of probabilities must be used in place of Table 34 above, since the weightings to be attached to the two variables in the predictive equation are no longer comparable to those of the period of 1923 to 1927. It was indicated above that in that early period high school percentile rank and the psychological test percentile rank were given equal weights, so that a C.A.R. could be computed simply by averaging the two percentiles. The recent research study, however, indicates that in predicting students' probable success in this college, high school percentile rank should be given a weight of a little more than two and one-half times the weight given to the total score on this psychological test.[5]

Again we point out the necessity in such research studies of

[5] Williamson, E. G., and E. S. Bordin, "Prediction of Success in the Arts College." To be published as part of a book summarizing prediction studies for all colleges in the University of Minnesota. University of Minnesota Press, 1941.

periodic revision of probability tables, such as are given above, through the derivation by statistical methods of the weights to be attached to the variables in the predicting of probable college grades of students.

TABLE 37

PREDICTION TABLE FOR DETERMINING PROBABILITY OF AVERAGE GRADE OF C OR HIGHER FOR VALUES OF WEIGHTED PERCENTILES. DERIVED FROM STUDY OF 1360 ARTS COLLEGE FRESHMEN IN 1938 [6]

Weighted percentile score	*Total number of cases*	*Number with average grade of C or higher*	*Per cent with average grade of C or higher*	
96-100	56	55	98.2	95.3
91- 95	92	86	93.5	
86- 90	94	74	78.7	
81- 85	145	101	69.7	72.2
76- 80	132	93	70.5	
71- 75	122	77	63.1	
66- 70	130	71	54.6	
61- 65	93	37	39.8	46.1
56- 60	98	31	31.6	
51- 55	86	28	32.6	
46- 50	70	21	30.0	
41- 45	81	15	18.5	25.5
36- 40	53	16	30.2	
31- 35	51	8	15.7	21.4
26- 30	33	10	30.3	
21- 25	15	2	13.3	
16- 20	6	2	33.3	
11- 15	2	..	..	16.7
6- 10	1	..	..	
1- 5	..	..	..	
Total	1360	727		

A new probability table was computed for freshmen enrolled in this same college in 1938. A comparison of the percentages given in the last column to the right in Table 37 with similar percentages in Table 34 reveals the fact that by the use of a new psychological test and new statistical weights we now are

[6] Williamson, E. G., and E. S. Bordin, *ibid.*

able to make approximately as accurate predictions of present-day students as Johnston made over fifteen years ago for students in this particular college.

It should be noted that there are fewer students with C.A.R.'s below 26 in Table 37 than appear in Table 34. This results in large part from the exclusion of most students with low aptitude ratings from admission to this college as a result of Johnston's early studies. At the present time applicants for admission with low C.A.R.'s are given additional tests and extensive interviews to determine whether other qualifications outweigh the probabilities for success derived from the C.A.R. As a result of this more selective admissions policy fewer students are admitted with low C.A.R.'s and these few are more likely to succeed than was true at the time of Johnston's investigations.

SUMMARY

If colleges are to discharge their responsibility to society for the careful training of students for the professions, then they must do one of two things: enroll all or nearly all who apply and then follow a rigorous policy of assigning failing grades to those students who do not live up to reasonable scholastic standards; or develop admissions procedures which will make possible the identification and rejection of students with low probabilities for success in class. The first alternative of admitting almost every applicant and then failing those who should not have been admitted is not a desirable practice in terms of the use of funds or the effects upon the morale of students. Years of research have yielded reasonably accurate methods of identifying students with low probabilities for success. A series of experiments in one college was described above in some detail to illustrate appropriate admissions procedures. The accuracy of results obtained by the use of selection methods perfected through experimentation has been demonstrated by follow-up studies which are in substantial agreement with results obtained in earlier investigations.

In the application of the results of these experiments, students with low probabilities of success may be refused admission to a college which seeks to maintain high standards of professional training. Such students, if identified early enough, may be advised to secure other types of training more in line with their level and type of scholastic aptitudes. In some cases this means that colleges themselves should set up new types of curricula adapted to the level of aptitude and types of interests of students not properly qualified for the traditional curriculum. In the case of other students the use of such selection methods permits the referment of students to training courses offered by non-collegiate vocational and commercial schools.

Part Eight

INDUSTRIAL PSYCHOLOGY

By

HARRISON MUSGRAVE
Southern Counties Gas Co. of Calif.
and
Southern California Gas Co.

CHAPTER 27

Selection of Personnel

There are almost as many methods used in the selection of personnel as there are people involved in doing it. It is therefore not surprising that the methods employed vary through all stages from the relatively adequate to the practically useless. By far the commonest method is the use of the personal interview, by which the personnel manager hopes to pick out well-qualified applicants for varied sorts of positions by talking across the desk to them for a half hour or longer. Certain information is gathered, concerning the type of previous employment, education, references, etc., but the personnel man's impression of the applicant usually appears to be the deciding factor as to whether or not he gets the job. This puts a great responsibility on the interviewer to try to understand and overcome his own personal biases and prejudices so that he can judge the applicant in accordance with his real ability to perform. That this is often impossible is attested by the stereotyped kind of employee hired by many executives.

Attempts at objective judgment of applicants. Some people who do hiring try to overcome this impressionistic effect by asking various objective questions designed to show whether an applicant has the necessary qualifications. Even here, however, it is very difficult to get away entirely from general impressions. Furthermore, an applicant automatically puts his best foot forward when applying for a job, and few indeed are the applicants who cannot maintain this good impression for a half hour at a time.

Many very inadequate methods have come to light. For example, one personnel man used to keep on his desk a box containing a few assorted nuts and bolts. Some time during the interview he would "accidentally" knock it onto the floor in

front of the applicant. As would be expected, the applicant would immediately lean down, pick up the spilled parts, put them back in the box and replace it on the desk. So far so good. But that was not enough. As soon as the box had been replaced, the personnel man would ask him how many pieces he had picked up. If he knew, his chances of being hired were pretty good, because "he was observant"; if he did not know, the interview would be terminated quickly.

Another executive, during the interview, would take out a cigarette and ask the applicant for a match. If the applicant immediately reached into one pocket, pulled out a box of matches and handed them over, he would probably be hired because "he had an orderly mind." But if he fumbled through several pockets before producing, his chances were slim. What happened if he didn't carry matches seems to be a moot point.

Pros and cons of the interview method. The pros and cons of the interview method for hiring have been discussed widely and there is no need to go into them thoroughly here. H. L. Hollingsworth of Columbia University published the most critical study of this technique in his book *Judging Human Character* (p. 62). He arranged for a group of people who were experienced in hiring to review the qualifications of a group of 60 applicants for a fairly responsible position which was actually open. Each personnel man was asked to interview all of the applicants and then rank them in order of desirability for the job. His hypothesis was that people who have had experience in judging men and who are customarily charged with the responsibility of doing so should agree fairly well as to the relative merits of the applicants.

A survey of the judgments showed marked inconsistency: one man received first, second, fifty-third and fifty-seventh places on the scale from various judges; the applicant receiving the highest average standing was ranked all the way from second to thirty-sixth place; and the two applicants who received the worst average rankings varied from sixth to fifty-fifth places. It appears from the results that using a group of interviewers to review the candidates had a greater tendency

to confuse the issue as to which one or two were the best than to narrow down the field to a few logical choices. Probably the diversity and indefiniteness of standards among interviewers were the major factors, along with the fact that very few organizations ever check the efficiency of their employees to see how well their methods of selection are working. This neglect of personnel audits permits the continuation of inadequate methods of selection.

In another case a large corporation tried the experiment of having all applicants interviewed by five trained interviewers, on the theory that this would tend to compensate for biases which one or two interviewers would be expected normally to possess. After a short time this venture was abandoned for it was found that only the most mediocre people managed to survive this fivefold test. An applicant with any signs of individuality was pretty sure to clash with one or another of the five interviewers and so be rejected.

This discussion is not meant to disparage unduly the use of personal interviews, for there are many situations in which they are both desirable and necessary. Especially is this true as a supplement to other more objective hiring programs. It is clear, however, that recent studies bring out the unreliability of the results attained by using this method alone. Probably most business men feel that they might just as well have a good-looking secretary as a homely one; but choosing a group of secretaries on the basis of appearance alone might be risky, for so far no one has adequately studied the relationship between good appearance and job efficiency. This might even be impossible, for ideas of what constitute good appearance are somewhat erratic.

It is true, however, that for years past many people have made honest and sincere efforts to devise more objective selection procedures which would find the right men for the right jobs. The devotees of phrenology, a theory exploded years ago, had their era of influence and many an applicant has gone to work because he had a strong chin or has lost out because of a sloping forehead. Many other will-o'-the-wisps have been

followed with equally indifferent success. More recently the use of psychological tests has come into prominence as an aid in selection, and it seems worth while to trace the origins and growth of one such program which has been carried along on a large scale.

The trade test. There are two general types of testing programs. The most common is the practical or trade test. In this the intention is to find out by a miniature performance or "work sample" how well the applicant has developed his abilities to perform certain clerical or mechanical tasks. This procedure, because of its very nature, has an intrinsic appeal and is rather readily accepted by management executives; for obviously if a person excels in the kind of work he is expected to do, he is the one to hire. There are, however, two objections to this procedure which are not immediately apparent. In the first place, the applicant's ability is being tested at a particular time, and there is no way of knowing whether someone else with greater latent capacity but less practice at the act in question might not in a very short time easily surpass the first applicant. In other words, because an applicant shows up best now does not mean that he will be the best for the job a year from now. And certainly if the company prefers to train its own employees, a test of present skill would be of less importance.

It appears, too, that the lack of actual ability to do the job is not the major factor in dismissals. A study made by the Bureau of Vocational Guidance of Harvard University and reported by Crane (p. 338) in his *Psychology Applied,* indicates that in over 4000 cases, 65.8 per cent of the dismissals were for lack of social understanding—insubordination, general unreliability, laziness, trouble-making, drinking, etc., and only 25.7 per cent were for incompetence. This should indicate that perhaps skill isn't the most important thing to test for in many lines of work. Failure to use skills possessed or inability to adjust to the social situation seem to be more important items.

Psychological tests. The other type of testing program deals with measurements of intelligence, temperamental traits,

etc., and only in some cases with clerical and mechanical ability as such.

It has been with great interest that the author has watched for several years the growth of the testing program instituted by Wadsworth,[1] in a large utility corporation. The program is of interest partly because of the logical and methodical steps which were followed in developing it, partly because of the methods that were used to check back on the efficacy of the various techniques tried, and partly because of the significant results which the system has produced.

By 1931 it appeared that results attained with the more usual employment methods, such as interviewing, letters of reference, etc., were not all that could be desired. It was decided, therefore, to look into psychological tests to see if employee selection could be improved. With this in view, the first step was to give intelligence tests to the entire existing force. There were two purposes: first, to determine if the score ranges found in different classifications of workers differed in any regular or usable sort of a way; second, to see if the scores of the successful employees in any given occupation differed from the scores of the less successful. To anticipate, it was found at once that the answer to both of these questions was "yes."

Employees were given two standard intelligence tests. As would be expected, the results of these two tests were almost the same. However, if the I.Q.'s from these two tests differed more than 5 points for a given person, additional tests were

[1] Humm, Doncaster G. and Guy W. Wadsworth, Jr., "The Humm-Wadsworth Temperament Scale," *Am. J. of Psychiatry,* Vol. 92, No. 1; July, 1935.

Wadsworth, Guy W., Jr., "Practical Employee Ratings," *The Personnel Journal,* Vol. 13, No. 5; February, 1935.

Wadsworth, Guy W., Jr., "Tests Prove Worth to a Utility," *The Personnel Journal,* Vol. 14, No. 5; November, 1935.

Wadsworth, Guy W., Jr., "How to Pick the Men You Want," *The Personnel Journal,* Vol. 14, No. 9; March, 1936.

Wadsworth, Guy W., Jr., "Temperament Tests as Personnel Aids," *The Personnel Journal,* Vol. 15, No. 9; March, 1937.

Wadsworth, Guy W., Jr., "Fit Employees to Their Jobs," *The Personnel Journal,* Vol. 16, No. 5; November, 1937.

given. When the test scores were tabulated on the basis of the occupation of the employees involved, it was immediately noted that there were definite similarities in the scores made by employees in the same types of work. Skilled workers, as would be expected, tended to show higher scores than unskilled workers. The distribution of scores for each occupation tended to approach the normal distribution. This, of course, answered the first question very neatly—that is, various occupational levels tended to be characterized by rather definite levels of intelligence.

Rating by supervisors. To answer the second question, as to the relation between occupational competence and general intelligence, it was necessary to get a measure of the relative value to the company of all the employees who had been tested. The best way to do this was to have employees rated by their immediate supervisors, the people who knew the most about them. Several types of rating forms were tried, including both standard published forms and some worked out in the personnel department. In general, the standard published forms did not prove very satisfactory, for supervisors seemed to resent having to express their ideas about an employee in terms of stereotyped phrases, or in terms of some point on a scale from one to ten for various listed qualities. Also, it was found in many cases that although an employee might be given a fairly high score on the rating form, the supervisor still didn't think much of him. Therefore a form was finally devised which allowed supervisors to express their answers to a few simple questions about the employee in their own words. Six questions such as the following were asked:

List the good points of this employee as you see them.

List the faults or limitations of this employee as you see them.

Considering actual ability shown on the job, would you hire this employee over again if you were to make the decision?

How do you rate this employee: outstanding, satisfactory, or a problem?

The following —— employees in your district are classified on the payroll the same as this employee: (List of Employees) If

you were to list these —— employees in order of their general value to the Company, as what number would you list this employee? Number ——.

One advantage of a form of this kind is that it forces the supervisor to think about the employee objectively, because it asks for facts about him rather than for an over-all impression. Also, after the supervisor has reviewed the various facts about each employee rated, he is in a much better position to render an objective decision as to whether the employee is outstanding, satisfactory or a problem. To avoid unfairness to an employee on the part of any one supervisor, each employee was rated by three of the supervisors over him. In this way, it was possible to get judgments from several of the people most interested in the efficiency of hiring. The man-to-man comparisons as to relative value to the company were compared with intelligence test scores. Coefficients of correlation between an individual supervisor's opinion of the rank order value of the men under him and intelligence test scores ranged from +.57 to +.87. The average coefficient for all such comparisons was +.68. These were correlations for many occupational groups, although for the most part the groups were fairly small. The interesting point is that the relative value a supervisor attached to his men was closely related to their relative intelligence levels, *even though supervisors had no knowledge at all of intelligence test results.*

This finding indicated that something was to be gained by following the procedure further: it appeared that tests might find very definite and measurable distinctions between successful and unsuccessful employees which could not easily be secured by other methods. As is true of most companies, this one had its usual number of problem employees who over the years had shifted from one job to another, usually in the direction of one suited to their intelligence level. Others had been left in their same jobs, more because it was hoped that they would finally succeed than because their lack of fitness went unrecognized. In many of these cases test scores showed they did not

have the necessary intelligence to handle their jobs adequately. This, of course, was an indication that test scores, if used in new hirings, might almost eliminate this type of problem in the future.

Minimum test requirements. Up to this point it had been determined, then, that various occupations tended to have their own characteristic test score ranges, and secondly, that the employees whom the supervisors considered to be the best tended to fall in the upper half of the test score range for each occupation. The next step was to turn this information around and to set up minimum test requirements for future hirings in each of the occupations so studied.

The median of the distribution of scores in each line of work was made the minimum for that occupation. Previous studies had brought out the fact that those scoring in the upper half in each occupation constituted the larger proportion of outstanding employees and the smaller proportion of problem employees. These minimum scores of course varied with the occupation, running as low as 90 I.Q. for some manual tasks, 100 for some kinds of clerical work, and higher for certain specialized occupations. They were reviewed every year and adjusted up or down in the light of information gleaned from the annual ratings turned in by supervisors. Figure 48 shows the results accomplished by this procedure. The total number of hirings included in the figure is 702. The people hired on a non-test basis number 594, and those hired under the testing procedure number 108. Even though in some cases tests other than those of general intelligence figured in the selection of test-hired employees, an acceptable level of intelligence was the most basic requirement in each case. For example, assuming the intelligence level to be adequate for two employees being considered for an opening, the one making the higher score in a mechanical or clerical test would be the one taken on. But the process was never reversed: an applicant scoring high in a practical test was not accepted if his I.Q. was below the minimum set up for that occupation.

In considering female clerical employees, an I.Q. of 105

turned out to be a good minimum. Scoring below 105 were 61 per cent of all the problem employees in this classification and only 27 per cent of the outstanding ones. Above 105 were found only 39 per cent of the problem employees and 57 per cent of the outstanding employees. The whole procedure assumes merely that by hiring within appropriate score levels, the chances of getting outstanding employees are increased and the chances of getting problem employees are decreased. With-

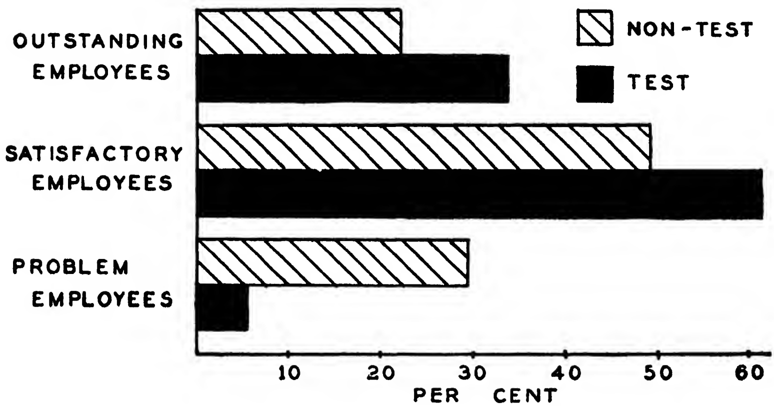


FIG. 48.

out doubt an occasional applicant, although he doesn't have the required intelligence level, would make an outstanding employee, but, on the average, the chances of this happening are poor. At any rate there is no question but that other methods of hiring also eliminate people who might turn out to be excellent employees. Interestingly enough, for certain kinds of work, maximum intelligence levels had to be used instead of minima. In other words, the higher the intelligence in this work, the more likely the individuals were to become problem cases. So, for these jobs, only applicants whose scores fell below rather than above a certain level were hired.

The Humm-Wadsworth Temperament Scale. The steps just reviewed had already considerably improved the quality of applicants hired, but further studies brought out another important factor in job success. As indicated in the study re-

ported by Crane, a surprisingly small proportion of dismissals occurred because employees could not do the work required at their job. Most workers were dismissed because they lacked social understanding or adjustment to their social situation. As more and more ratings were examined in the present studies, it became clear that most of the criticisms of employees were not about the actual work performed, but rather about matters of personality. Supervisors would say that an employee was conceited, or that he couldn't be told anything, that he could do the work but sometimes didn't seem to want to, or that he would "go great guns, but you had to build a fire under him every now and then." Situations were uncovered where one supervisor would think a man was "loud-mouthed," talked too much, and was egotistical, whereas another supervisor would consider the same man to have lots of good ideas, be a good conversationalist, and get along well with people. It therefore seemed logical to try to find some way to match people who had to work together so that they would get along well. Various standard personality tests were experimented with, but none of them seemed to be broad enough to accomplish the desired results. It was then decided to try to work out a new test which would at least in some measure accomplish this end. Mr. Wadsworth, in collaboration with Dr. Doncaster G. Humm, spent some time in devising and standardizing the Humm-Wadsworth Temperament Scale, the development of which has been fully reported elsewhere. The careful and judicious use of this scale enables one to determine, in a large percentage of the cases examined, several temperamental traits of the individual. Some examples of these traits are:

1. **Self-control.** Everyone loses his temper from time to time, but some take offense and are ready to fight, whereas others can control this impulse and work out differences of opinion in a socially more acceptable manner. Such factors of control can be measured quite reliably, and obviously are a desirable thing to know about a prospective applicant or employee.

2. **Egocentricities.** These have to do in general with the idea of selfishness, not so much in the narrow sense of wanting the biggest share of things as in the broader sense of always looking after one's own interests in doing or not doing anything. Some people are always on the lookout to see "what there is in it for them" before they will move a finger. Obviously we are all motivated by this feeling to a certain extent, but when it gets out of bounds it can be a very disturbing and disrupting influence among employees on the job. Tendencies of this kind can also be uncovered by appropriate measuring methods.

3. **Emotionality.** Many people have very strong likes and dislikes about those with whom they are associated, or about what they are doing, whereas others are quite placid and unemotional about what goes on around them. People with marked ups and downs of emotion often have to be handled very carefully, and can cause much confusion around an office when they go on an emotional spree. Moroseness and general uneasiness among employees is likewise disrupting. On the other hand, if the emotional type of person can be held under reasonable control, or can so hold himself, he can be used to great advantage in sales and other public contact work, since for the most part he truly enjoys being with people and getting along with them. His enthusiasm is infectious and can often go a long way in keeping up the feeling tone and spirits of an office force. At any rate, emotionality can be measured.

4. **Introvert tendencies.** These tendencies are exhibited by shy people and sometimes by those who are conceited, suspicious or arrogant. Introverts tend to be good at routine or detail clerical jobs, since they like to work alone and don't mind going along for days on end with little or no social contact. Since such tendencies also can be measured, it should be obvious that a knowledge of the general characteristics of the people with whom one has to deal would be helpful in making proper placements or transfers. This knowledge will help not only to fit the right temperaments together, but also to place them in appropriate kinds of work. There is, of course, considerable confusion in definitions of the various temperamental

traits as they are used by different investigators, but from the practical point of view it is especially important to know what one is talking about in terms of the everyday descriptions of employees as given by their supervisors.

Practicality of temperament tests. Practical tests and tests of intelligence for the most part give us a quantitative measure of people's abilities and traits. No such claim can be made for the tests of temperament. It should rather be said that a test of temperament gives an excellent clue to what kind of behavior to expect in a practical situation. The use of such tests should always be supplemented by a clinical study of the actual behavior of the person involved. As an example of this, it might be pointed out that a person who shows up as extremely paranoid on a temperament test may display this characteristic in one of two ways on the job. He may be argumentative, unwilling to listen to others, and hard to get along with, which of course would be undesirable. Or he may display the paranoid trait as a driving force which causes him to be tenacious and carry through to the end. Even when he meets rebuffs and obstacles he may keep on trying in one way or another until he finally accomplishes his objective. This of course would be a desirable attribute. It is only by clinical observation that one can determine which way the paranoid trait operates in the individual in question. On the other hand, it also is true that in many cases it pays to use the scores made on a temperament scale directly, for if an applicant shows up with a series of temperamental traits which are very similar to traits which have often been characteristic of problem employees on the job, it is hardly safe to take a chance on him on the theory that he might be the exception.

Additional information obtained. Assuming that scores on personality tests should be used mainly as clues for further clinical investigation of applicants, the question then arises as to where such additional information can be obtained; for clinical case histories are generally not available to personnel men. Extensive studies of letters of recommendation and the questioning of neighbors, relatives and friends have indicated

that very rarely indeed does a personnel man find out anything unfavorable about the applicant. This is true undoubtedly because no one wants to stand in the way of somebody else's getting a job. One quite successful technique has utilized the services of an outside investigating organization. In this case a national credit service was used. Its investigators not only got routine credit information, but questioned former employers and persons given as references on various points of temperamental significance. In a large proportion of cases it was demonstrated that the information so turned up fitted in very nicely with the scores made on the temperament scale.

The temperament tests and incompatibility. The temperament scale has also been found very useful in carrying on case studies of incompatibilities between the supervisor and employee or between two employees. As mentioned before, the work an employee does is a relatively small part of the supervisor's over-all impression of the employee's value to the company. A supervisor who likes an employee will overlook many of the technical faults in his work, and, if the liking is absent, will tend to minimize even excellent job performance. The social situation cannot be separated from the job situation. And, in checking up on cases of this nature, the cues given by the temperament scale are an aid in knowing what to look for in the relationship between the two persons involved. Test results are also an aid in knowing how to approach the individual when talking to him in an attempt to overcome personal difficulties. In 200 case studies the supervisor's ratings indicated that 120 had definitely improved and 40 more had improved to a certain extent as a result of a frank talk in the personnel department concerning what was wrong with their actions in connection with other employees.

The interpretation of results obtained by the testing technique as here presented might appear too optimistic. Certainly no such claims are intended. A personnel department does not seek to hire a group of super-applicants, but rather people who have the desirable characteristics of the best workers already employed, and not those who seem to resemble the problem

cases on the payroll. In general, when there is an opening, the personnel department tries to find three or four applicants whose test scores indicate that they could perform satisfactorily on the job, and then have the supervisor pick out from these candidates the one with whom he thinks he can work best. The results of this procedure have been a very distinct improvement over results obtained by other means. And it must be remembered that the people who finally evaluate such a testing program are the supervisors who by their ratings tell the personnel department how the applicant is getting along on the job from year to year.

Current Trends in Employee Training

METHODS of employee training have, in general, been developed to fit the particular needs of the business or governmental agency that develops them, with but little reference to what other such programs are doing. For this reason, it is difficult to factor out and describe "trends." Perhaps the best approach is to generalize briefly about employee, or "in-service,"[1] training, following up with a few typical examples of present-day programs as illustrations. Three such trends seem of particular interest.

In-service programs. First, there is a gradual but steady increase of in-service and employee training programs, which tend to include supervisor training, as well as training in an ever-widening range of vocations. For example, the Los Angeles Playground Department has recently extended its conference series to include occupational groups such as laborers, mechanics, etc. They have three objectives: to train new employees, to give older employees a chance to learn about and prepare for the next job ahead and to learn how to do their present jobs better. Janitors, for instance, have had conferences on better methods for sweeping floors—with beneficial results! At the other end of the scale, a large electrical equipment company has instituted a series of training conferences for their foremen. This includes 21 men, all the foremen up to the general manager, and is designed to promote improved human re-

[1] The terms "employee" and "in-service" training can really be used interchangeably. However, "employee" is usually used in referring to private business, and "in-service" when discussing governmental or other public servant training programs. As the book goes to press national defense committees are getting under way to provide adequate methods of training skilled workmen in defense industries which in many localities have already exhausted the supply of available trained workers.

lationships between supervisors and those whom they supervise.

The conference method. The second trend in employee training is a trend away from the lecture, textbook, and other rather formal training procedures, to the conference method. This probably is not the fastest method. It undoubtedly does not cover as much ground as could be covered in other ways, but it appears from experience that the most people get the most good from it. The lecture method is the easiest. A department head decides his group needs a little material on building customer relations, so he arranges and holds a dinner meeting for his employees. He hires a good speaker on customer relations to come in and give them an hour's talk. When he is finished, the executives present have seen that their people were there, they have heard the speaker say the right things, and can clap their hands and say, "There, we have taken care of customer relations; now let's go on to something else." But how do they know that any of those present were actually listening, and, if they were listening, that they really appreciated what was being offered them? And even if they did appreciate and understand what was being offered, will they translate this into actually altering the procedures they use on the job? These assumptions are open to question. Probably everyone has had the experience of reading a book on the order, say, of Dale Carnegie's work on influencing people. The average person is apt to agree that he offers some ideas that would be helpful, if practiced. The next day he may tell a friend, "I read a good book by a man named Carnegie last night on how to influence people. I sure am going to try it out." A week later he meets another friend, and says, "Gee, Bill, I read a swell book last week about getting along with people by some fellow—I can't think of his name right now, but you sure ought to read it; it's good." And three weeks later, he has forgotten there was such a book.

Mere exposure to things is no guarantee that we learn them. As an example, think of two main streets about a mile apart outside of the central business district of some metropolitan

area, both of which you have crossed frequently in your normal comings and goings. Can you name the streets, in order, that lie between them? You may have been exposed to them a thousand times. It takes more than exposure to insure learning, and it takes more than the fact that you have learned something to insure that you will put it into practice. It is necessary to develop habits of doing the thing in the right way, and the conference method, in the field of ideas, comes closer as a practical method of forming such habits than any other way so far developed.

Efficiency of training programs. The third trend is the most important, but it is hard to describe in a sentence. It has to do with the efficiency of training programs. In-service training directors generally are making a concerted effort to design their programs so that they actually produce results, rather than just expose people to good ideas. Obviously, the object of employee training, as well as any other training, is to make it possible for those involved to learn something. This leads to the question: How do we learn most effectively? Regardless of how we approach the problem of learning, two factors seem to stand out. These are that learning is a function of the whole organism rather than of a few isolated parts of it, and that we learn by doing.

If it is not immediately apparent that learning is a function of the whole organism, watch people taking their first typewriting lessons. Not only are their hands working hard, but their legs and feet are all tense, their faces screwed up, and, in general, the whole body is in a state of considerable activity, most of which is of doubtful value. Gradually as they improve, these various unnecessary bodily activities drop away, and only those parts of the body immediately involved in typing appear to be accomplishing all the work while the rest appears relaxed. But even then, it cannot be said that only the hands and arms are involved in the typing process. The rest of the body has to arrange itself in an appropriate position so that the hands and arms can function most effectively.

If the whole body is involved in learning, as it is in any

activity we undertake, we might reverse the idea and postulate that the more parts of the organism we can bring into play in effecting learning, the better off we are. As an example of this, recall that taking notes on a lecture, even if they are thrown away at the end of it, is a help in the learning process. The material must be both heard and considered before it can be briefed into notes. Secondly, the muscles (and kinesthetic sense) are called into play in writing it down; and, thirdly, the visual sense then offers a still different exposure. Thus, instead of merely hearing it, the material is presented through at least three avenues, which insures a greater chance for retention.

We learn by doing. Our second point above was that we learn by doing. To demonstrate this, those of you who are golfers have undoubtedly learned that you cannot go to the movies, watch a few exhibitions on the screen of perfect shots by Bobby Jones, and then go out and do likewise. It takes a lot of hard work and practice to play even moderately well.

These two points will serve as a basis for understanding what appears to be happening in present attempts to work out more effective methods of training. First, the more modalities of the man that can be called into play, and, second, the more he can be made to participate himself in what he has to learn, the better will be the results.

Let us now examine the application of these two principles by a brief résumé of five specific training courses. Each of these five courses illustrates certain relatively new departures from older methods used in this field. For comparison we will first describe the traditional absorption method.

The absorption method. In the past, a great deal, if not most, of employee training has been carried on by the absorption method. This method is essentially simple. The employee reports to his boss the first day he comes on the job and is turned over to someone who is doing the kind of work the new employee has been hired to perform. The older employee introduces him to some of his fellow workers and shows him around a little bit, spends two or three hours telling him what

he is supposed to do and then leaves him to carry on, with the remark that if there are any questions, he will be glad to answer them. The new employee starts out and every time he gets stuck he bothers someone to know what to do. About noon, the supervisor comes along and asks how he is getting along. The answer is usually "fine." In the middle of the afternoon the older employee checks back again, asks the same question and probably gets the same answer. After two or three days of this kind of procedure, if the employee is not making too many mistakes, he is considered to know his job and is left to his own devices.

A frequent variation of this procedure is to put the new employee with several older employees doing the same kind of work for a day at a time in the vague hope that in this way he will learn a little bit from each one and so get a more rounded picture of his duties. However, there is no assurance that the new employee won't pick up many of the bad habits of each of the older employees, as well as some of the good ones. More and more it appears that both business and public service are finding it not only economical but also necessary to devise and execute a more definite training program for new employees entering their service.

1. **A ten-week training method.** Let us look at a more modern method. An excellent example is exhibited by a large utility corporation which has a sizable force of what might be termed roving mechanics. They go out on call with their own trucks and repair various kinds of mechanical equipment which have been reported to be out of order. In the past they were instructed largely by the absorption method outlined above. Recently, however, a ten-week training program was instituted which covers very nicely the basic points outlined earlier; namely, learning through a number of sensory fields, and learning by doing. The first day on the job they spend the morning riding with an experienced man, observing what he does and getting a general idea of the kind of work that they will be called upon to do in the future. In the afternoon they go to a training class where they are introduced to the staff,

told about the kinds of things that will be expected of them and about the company organization, and are given operating instructions and various manuals covering the work in question.

The second day and for the remainder of the week, the newcomer is again sent out with an older employee to observe specifically the kind of work that is being done. He is not allowed to do any of the actual work himself, but is required to help on odd jobs, such as wiping off the car, carrying parts, and is supposed to learn the location of stationery supplies, first aid equipment, and, in general, observe as much as possible.

The second week he again spends half a day in classroom training. He *is quizzed on what he has learned* and is *trained specifically* in the exact way to do the simplest kind of job he will later be expected to perform. For the balance of the afternoon he is given an illustrated lecture covering the over-all aspects of the company's operations, where it gets its supplies, how it operates, etc. The other four days of the week he again goes out as a student helper with an older man, and actually performs those duties in which he has received training.

Each week thereafter he receives specific training in a small number of duties which are slightly more complicated than those he has had previously, and then goes out and actually performs all the duties that happen to arise on the route which he is covering, under the supervision of the older men. To state it differently, he is led by easy steps from the simple to the complex, the new things he learns always being extensions of the simpler things that have gone before. In the classroom the training consists of several parts. He is shown slides of various kinds of equipment with which he will be required to deal, and the parts and operation of this equipment are described to him. Then he is shown cutaway pictures and line drawings, and has the internal workings described. Following this, he is introduced to the actual equipment in the classroom and its operations are again gone over. The next step is to have him manipulate the equipment himself, aided by the instructor.

Thus we note that he sees it, hears about it, and actually handles and works with it.

This same general process carries on for a period of ten weeks, at the end of which time he has covered the major portion of his regular duties. During this ten-week period the various older employees with whom he rides turn in a weekly report on his progress. He also is given a weekly quiz on what he has already done, his progress is watched closely, and, if he seems weak in any places, special emphasis is placed on these in the classroom and in the field.

During the tenth week he takes a final examination and undergoes a full day's field check, doing all the work called for under the watchful eyes of an instructor. The instructor offers neither comment nor help unless something is done, or left undone, in such a way that the people being served would suffer. The instructor turns in his report, and on the basis of all the available facts the new employee is either dropped or released to the operating department for a regular job. Incidentally, during this ten-week period he is on the payroll as a regular employee.

2. **Continuous training.** In the foregoing description the training of *new* employees is considered as a function all by itself. It does not mean, however, the employee is considered fully trained at the end of the ten-week course. As a regular employee, he engages in a second training program. Since he is actually on the job, this program is not, of course, quite so intensive. He is now working regularly, driving a truck of his own. However, he comes in for a two-hour classroom session approximately twice a month. The classroom is equipped with cutaway models of household appliances, descriptive material, a blackboard, and other items necessary to cover the desired ground adequately.

During these two-hour sessions his training embraces descriptions of the more complicated and unusual kinds of equipment he is apt to encounter. Also, he is continually brought up to date concerning changes in new models, entirely new devices which may come on the market, changes in operating proce-

dure, and anything else that he should know. This process is a continuing one and lasts as long as he is on the job. In general, the procedure is to describe a new device such as a refrigerator or stove, accompanied by slide pictures thrown on the screen. The employee is then shown the actual apparatus and is shown how to disassemble, reassemble and adjust it for any operation that might normally come up in the field. After being shown how to do it, he is required to do it himself in the presence of the instructor and is given suggestions concerning special techniques or short-cuts that might come in handy.

Note here again that the process involves hearing, seeing, and actually doing. As a further step in the training procedure, the foremen occasionally visit locations where the man has completed an order and recheck the work he has done, not with the idea of finding fault or being critical, but rather to see that the work has been done in an appropriate and efficient manner. They also seek to determine whether the man needs further training in any aspects of his work, or whether any suggestions could be offered that might be helpful.

The men, of course, know that these field inspections are being made from time to time, but do not know exactly when.

This general plan, you will note, tries to insure the best possible training and also serves as a check on how effective the training actually is in each individual case. Very often, as a result of these checks by the foremen, the training methods or techniques are adjusted in some particulars to take care of errors or omissions that appear in the work of more than one man, which, of course, is apt to be an indication that it was the training method and not the negligence of one man that caused the error.

(Note this last point. Checking the man's work is done in large measure to see if the *training program itself* is adequate. This idea of evaluating the results of training as a measure of the effectiveness of the program might be considered a fourth trend in the field. However, in the author's experience, the practice is not yet sufficiently widespread to be called a trend

The tendency is still to blame the man for not learning, rather than to examine the teaching techniques *after they have once been decided upon and set up.*)

3. **A conference method.** There is a third type of training which has some features that, as far as the writer knows, are used by only two companies in the country. This program is carried on under the pure conference method and has behind it three basic theories: first, as mentioned before, that we learn by doing; second, that everybody has at least a few ideas about almost any subject that might come up (this, of course, would be especially true with respect to the job they are doing and the general operations of the company for which they work); and, third, that we do our best work when we are not under any strain and are free to say what we like. Translated, this third point means that employees of the same rank meet together in groups of about ten with no one of a higher rank present at the meeting at any time.

The first step in setting up the program is to give some employees rather intensive training in the art of acting as a conference leader. They are impressed with the fact that they are not expected to be teachers or instructors in any sense of the word, but are merely expected to keep the group in order and to keep them discussing and developing ideas about the topic the group itself decides to discuss. Except for a manual on the technique of leading a conference, which is given to the leaders, the program has no outline, stated topics for discussion, or anything else to guide what goes on in the group conferences. The object of the course is to improve customer relations. The first time the group meets, the leader, who has been trained, explains to them why they are meeting—that is, with a view to improving customer relations—and tells them how the program is to operate. He then asks each member of the group to write down and hand in what he calls a work experience. This is nothing more nor less than some work situation in which the employee has actually found himself. In general, these situations should be rather simple. An example might be: An order clerk reports that a customer came

to the window to sign up for gas service. In the process, she was asked to put up a deposit and she asked, "Why do I have to put up a deposit? The department stores don't ask for them when I open a charge account."

The work experience as turned in does not include what the employee told the customer; it is merely a situation that has occurred. The group members discuss the situation and decide among themselves what would be the simplest and most effective answer to the question from the customer's point of view. They carry it one step further and even work out the exact language to be used in answering the customer's question. The point is that it is one thing to decide to be courteous to a customer, but it is another thing to decide just what to say that actually is courteous. And, in general, after going through and working out the answers in terms of these exact words to be used in answering a large number of common situations, the members of the groups are in an excellent position to use similar and good answers on the job, as like situations arise, because they themselves have gone through the process of working out effective answers.

After each meeting the group turns in a summary to the program leader, briefly describing the work experience discussed, the factors of interest or importance about it which the members of the group brought up, and the specific answers which the group decided upon as the best to use in that situation. Provision also is made on the summary sheet for questions of any sort which the group wishes to ask about company operations or policies. You can see from this that it is a purely anonymous procedure and that no one in the management ever knows what any given employee might have said.

After two or three questions have been turned in and answered, and no member of the group has been discouraged from asking such questions, they soon learn that the system is really designed to help them, and the variety of questions that come in is astounding. It is also astounding to find how many questions come in that the management normally expects every employee to know. This demonstrates again that the manage-

ment has not in the past taken enough time and trouble in telling employees things they wished them to know.

One sidelight on this program has always been amusing. In describing the plan to supervisors, who, of course, are not included, they are told what it is all about and what might be gained from it. After a little discussion they readily agree that we learn by doing, and that everybody has some ideas about the things on which they are working. But when it is pointed out that the groups will meet entirely unsupervised, there is always a hush. Invariably one of the supervisors will finally break down and say, "But, with them in there all by themselves for an hour, they are going to talk about us"; and the answer is, "Yes, they probably will, but they talk about you anyway!"

Furthermore it is rather difficult for any group of ten people to get together and criticize some one person for an hour. If they do start out, they tire of it pretty soon and go on to something else. The other point that is sure to come up is, "Well, they are going to waste an awful lot of time." The answer to that is, "Well, they are going to waste a little time, but if you want to see some really 14-karat time wasting, go to a directors' meeting some time." Obviously, these two comments are somewhat flippant, but the two points are hard to answer, and after the program gets under way the results speak for themselves. As a matter of fact, having in the past several years gone over 5000 summaries turned in by discussion groups, the writer has yet to come across one that was out of line in any sense of the word, or one that was just plain carping criticism. Employees for the most part appreciate an opportunity to get together and discuss the problems of their own jobs among themselves. Incidentally, a great many valuable suggestions and constructive criticisms have come through on these summary sheets.

In some cases this customer relations aspect of the program has been varied, and instead there have been introduced manuals of company information, sales points, or other facts, which are given to the leader for reference, and em-

ployees under the same general method try to work out the best answer they can for sales arguments, or why things are done the way they are. The manual is there for reference or amplification of the points they bring out themselves, when and if it is needed. It is really surprising how much interest this method of training evokes from the employees involved.

4. **The periodic report method.** A fourth method of employee training, though not exactly new, is practiced far less than it deserves. Have you ever had your employer or your teacher come around and tell you just where you stood as far as your work is concerned? It is all too infrequent. If you do something wrong, or get in trouble, you are sure to hear about it. If on rare occasions you do an outstanding piece of work, you are apt to get a little commendation for it, but if, as is true of most, you are an average satisfactory employee, or student, you probably never hear anything. To most of us this never-hearing-anything is worse than being "bawled out" once in a while, and leaves us in a constant state of wonderment as to how we stand and what is to happen next.

It would be an excellent idea, from the standpoint of morale, if every employee and student were talked to by his supervisor or teacher at least twice a year. The supervisor or teacher should know enough about the quality of the work being done to give a *factual* account in terms of the work performed, as to how the person is getting along, what he is doing right, and what isn't so good. Included in this should be a definite offer of help in anything the employee or student needs, or needs to know. If more training is indicated, there should be a specific attempt made to see that he gets it. There are a few companies and universities that have a definite policy of doing this regularly. Where it has been honestly attempted, the recipients have been found to be highly appreciative.

5. **The training of supervisors.** It is interesting to note that of all the talk and activity in the employee training field, 95 per cent is directed toward the rank and file employee. How about supervisors? Why is it that, if training is so essential, they are usually entirely left out?

As a partial answer to this question, a large corporation, in the early part of 1940, suddenly decided it would be well to institute a training program for some of their higher ranking supervisors. The conference method was chosen. To the training director was delegated the responsibility of being the group leader. This rather disturbed him at first, for two reasons. In the first place, supervisors as a group do not approve very much of being "trained." After all, they are the ones who are paid to know the answers and to tell others what to do. (Probably the higher one goes in the supervisory ranks, the truer this becomes.) In the second place, the director felt he did not know much about training supervisors himself.

After considerable thought, and perhaps a little worry about this problem, it was solved by adopting some advice from the writings of H. C. Link. Why waste time and effort trying to *overcome* obstacles? Why not, instead, *avoid* them? Acting on this suggestion, it was decided to let this group of supervisors train themselves without knowing it by developing a training course for lower ranking supervisors under them.

Obviously they could not develop such a program without going through it themselves. And, in passing, you might note here the strong use of the principle of learning by doing.

Accordingly, at the first meeting, the idea was broached and accepted with considerable enthusiasm. The general nature of the project was discussed, as well as the aims to be accomplished, and a steering committee of three was appointed to meet immediately and decide specifically what steps should be taken up and in what order.

The steering committee met in due course and decided it would be most advantageous if the group proceeded along the following general lines:

1. To determine all of the fundamental duties and functions that should be the responsibility of every supervisor. (These fundamental functions and duties should presumably vary only in degree between supervisors of different rank.)

2. To develop specific methods for accomplishing the duties and functions agreed upon under (1) above.

3. To illustrate and present the specific methods under (2) above by using actual cases and examples.
4. To summarize in written form the material developed in the above studies, possibly even in the form of a supervisory manual.

The program as thus outlined moved along smoothly and effectively. It had the advantages of accomplishing the original objective of giving the supervisors involved definite training along the lines of supervision, and the further advantage of combining and integrating the ideas and knowledge of twelve experienced supervisors into a standard, effective form of procedure for training others in the future.

As a sidelight, it is interesting to note that these men enjoyed the task they set for themselves, and that all of them did a considerable amount of work between the conference meetings in preparation for the next one.

These case histories are, of course, only examples of many things that are being tried. However, they do show some general directions in which progress is being made.

We are now going through a marked expansion in the field of training. This has been produced, obviously, by the accelerating demand for skilled workers in industry due to the National Defense Program. Little or no training was carried on in the '30's, for in general there was an ample supply of skilled labor for the available jobs. However, advancing age, retirements, deaths, loss of skill through disuse, and other causes have all operated to curtail the supply existing in 1930, and the moment expansion in production started the pinch began to be felt. A forecast of the number of skilled men that would be needed, and a survey of the training facilities existing, indicated clearly that further activity in the training field was needed on a large scale.

Late in 1940 the Labor Division of the Advisory Commission to the Council of National Defense set up a "Training Within Industry" section. In turn, 22 Districts were set up, each with a local Director, advisors from business and labor, and a panel of experienced training men to act as consultants.

These men were all loaned for this work by their respective companies. Early in 1941 this whole group was taken over by the Office of Production Management and this in turn was taken over and expanded by the War Production Board in January, 1942. At the present time panel members are giving concentrated ten-hour courses on supervisory training and job instruction. These have been carefully worked out and have proven very effective.

The objective of this program was twofold:

1. To urge defense contractors to train their own employees, and
2. To furnish them technical assistance in how to do such training.

These small groups of consultants cannot hope to do much training themselves, of course. Their function is to survey training needs in the different plants, and then to get management together with the appropriate city, county, state, and federal agencies which actually carry on whatever training programs are needed. There has been a high degree of cooperation evidenced all along the line, and the cases are rare indeed when management does not take advantage of the opportunities offered, or when the public agencies are unable to work out some means for being of real service. Services available cover every phase from pre-employment right up the line to supervision and executive training.

References

1. Greene, J. H., *Organized Training in Business,* Harper, New York, revised 1937.
2. Peffer, Nathaniel, *Educational Experiments in Industry,* Macmillan, New York, 1932.
3. Beckman, R. O., *How to Train Supervisors,* Harper, New York, 1940.
4. Cushman, Frank, *Foremanship and Supervision,* Wiley, New York, Second Edition 1938.

5. Walton, Albert, *The New Techniques for Supervisors and Foremen,* McGraw-Hill, New York, 1940.
6. Halsey, George D., *How to be a Leader,* Harper, New York, 1938.
7. Leigh, Robert D., *Group Leadership,* Norton, New York, 1936.

Consumer Surveys

DR. GALLUP'S Institute of Public Opinion, the Fortune-Roper Surveys, and others have recently brought this field of applied psychology into national prominence. It is interesting to note that as short a time ago as 1930 such surveys were practically unknown. Currently, the survey technique is being widely used by business to answer such questions as: "How do our customers like and use the products they have purchased from us?" "What do they think we should do to improve them and make them more usable?" "How are our sales-people presenting our story?" "What is our most effective type of advertising?" "What are the buying trends among the different brands of the same type of article?" "What is our reputation and standing with the buying public?" and many others. That this whole survey field should have taken such a strong upturn in a short time may appear strange, since it is not directly productive of profits. However, even before 1930 there was a readiness and a demand for methods of accurately appraising buyer's feelings and ideas on many counts.

At present, business is in the throes of making a very interesting and significant change in its general method of approaching and selling to customers. It is returning, we might say, to a more primitive underlying idea, but developing it on a sound and usable basis. The first rumblings of this trend appeared in the middle twenties as a reaction to too many so-called "high pressure" tactics. The trend was not really felt, however, until the recent depression was well under way, and "high pressure" methods were more annoying to the public than satisfactory in their results to business.

Early selling methods. Perhaps a brief glance at the origins of selling will make this point clearer. A few centuries ago,

salesmanship was practically unknown in the modern sense of the word. It arose in the first place as a result of the demand for certain simple goods and commodities. The point to be brought out is that the first salesmen did not really *sell,* they merely carried with them from place to place *things that people wanted, or needed.* More often than not, they merely traded items. As the number of things to be traded grew, it became impracticable for one individual to handle them all, so there was a gradual splitting up of goods into types; and thus the organization of guilds and crafts was started.

Selling as an end in itself. These guilds and crafts soon reached a point where division of labor was necessary and inevitable, so some individuals specialized in making the objects and others in trading and selling. As soon as this happened, selling automatically became *an end in itself.* It began to be necessary to sell what the workers made, instead of having workers fill separate orders for each individual buyer. At this point it was natural for salesmen and those in charge of selling to look for and try to develop methods for more efficient selling. The question was always, "How can we put more effectiveness into our selling attempts?" Efforts were directed towards beating down the customer's mythical "sales resistance." This was not a surprising attitude; rather, it was one to be expected under this method of growth. The interest and business of these men was selling, and that was the result. In fact, they finally became so proficient at it that they overshot their mark considerably. As evidence of this, note that the idea of high pressure salesmanship is not in very good repute at present.

The high pressure method. High pressure selling probably reached its peak during the boom of the 1920's. No one bothered much about it then, for people were in a buying mood. However, when it carried over into the depression years, and we had to curb our expenditures, we began to resent this type of approach on the part of the salesman. So now we find, as mentioned before, the cycle is being completed, and salesmanship is rediscovering the notion *of trying to offer people what*

they want, rather than forcing on them what salesmen happen to have for sale.

Selling customers what they want. That business actually is adopting this point of view hardly needs verification. We need merely to cite, as one outstanding example, the work of Henry G. Weaver, of the Customer Research Department of General Motors, which has sent out millions of booklets to people all over the United States, picturing various types of wheels, motors, fenders, etc., and requesting people to check their preferences and return the booklets.

This same trend and feeling has made itself felt, perhaps even more noticeably, in other ways. A score or more of recent books all emphasize the same note. Among the better known of these are: Carnegie's *How to Win Friends and Influence People,* Morgan and Webb's *Strategy in Handling People,* and H. C. Link's *The New Psychology of Selling and Advertising.* Since historians agree that current literature is a good mirror of popular sentiment, we are probably safe in concluding that this relatively new group of books is a significant indication of the scope of the present trend toward leading people rather than forcing them.

Avoiding sales resistance. Better than anyone else, Link has embodied the whole spirit of this trend in one phrase in describing the content of his book. His descriptive phrase is "How to *Avoid* Sales Resistance." This symbolizes the whole idea of finding out what people want and offering it to them in the way they would like to have it. There is an old saying that the best way to keep dogs from chasing cats is to remove the cats. Similarly, overcoming sales resistance should best be accomplished by finding out what people want, and appealing to their needs and desires in such a way that resistance and objections never arise.

So much for generalities. Let us assume, for the sake of argument, that such a new trend is actually in progress. Then the question logically arises, "How may we determine people's needs, attitudes, ideas, and desires?" To this, the only safe answer is to go to the people themselves and ask them; not to

one or two people, but to large numbers of them. Everyone has his own individual ideas and his own individual interpretation of what is going on. The chance that any one person's ideas are representative of the whole community is almost infinitely small. This is just as true of the executive in his office as it is of the man on the street. To know what people think and how they feel about things, it is necessary to pool the opinions of an adequate sampling of them, chosen at random from the community in question.

The survey method. To this end survey work of one kind or another began to appear about 1930, in an effort to determine what people wanted and how they felt on certain questions. True, surveys and questionnaires have been used for years. But there had been few scientific attempts made to determine how accurate the samplings actually were, how many samplings were necessary for a desired degree of accuracy, the relative merits of written questionnaires and oral interviews, the effect on the results of changing the wording of the questions, and many similar problems. The work of the Psychological Corporation in New York has been outstanding along these lines. On the basis of their work, and that of many other groups, the sampling required and the statistical aspects of survey work can now, for all practical purposes, be considered as resting on a fairly sound basis.

A typical survey experiment. Let us examine a typical example of the use of this method and some of the techniques involved.[1] In a large gas utility company, rumors filtered up to the executives in charge of sales that the company was losing gas range business because people were being informed that the natural gas supply was rapidly becoming exhausted.

Such an argument had not appeared in print, either in the form of publicity or advertising. It seemed, therefore, that this idea was either an easy excuse used to explain the loss of a few sales, or, if it had any foundation, it must be coming verbally from salesmen handling competing lines of merchandise.

[1] From unpublished studies by the author.

There was no way of telling which of these or other alternatives was the right one; nor could it be told, if there was some foundation for the report, how widespread the idea might be. If it did have any foundation, it would certainly be worthy of consideration and, possibly, a counter-attack. Such a counter-attack would be expensive, and it seemed wise to find out what the true situation actually was.

Following the assumption just mentioned, that if the permanence of the natural gas supply was being questioned it must be through the influence of salesmen, it was decided to survey a large number of salesmen handling gas and other cooking appliances to see what information they were giving customers, and what their selling arguments were.

To do this, four married women were selected as interviewers, all of whom had had considerable previous experience in this type of work. They were given twenty-five addresses apiece, and materials for reporting the interviews. They were told to act like bona fide buyers of a fairly good range, but to have no preference for gas or other fuels.

The 100 stores included 13 gas company offices, 11 electric power company offices, 29 gas range dealers, 20 electric range dealers, and 27 dealers handling both gas and electric ranges. Among the dealers were twenty large department and furniture stores handling ranges.

An additional procedure was introduced in this survey. It was felt desirable to find out what salesmen normally told customers, and also what they would say if pressed to work hard for the sale. Therefore, in those places where appliances were actually sold, the interviewers were instructed to make two calls, at least a day apart. On the first call the interviewers were to be "passive buyers," that is, to be interested in and listen to everything the salesman had to say, but to ask no questions about the relative merits of gas versus other fuels. When the salesman "ran down," the interviewers excused themselves, saying they would be back. Immediately upon leaving they wrote down on the blanks provided every argument the salesman had offered them, in his own words when-

ever possible. In this way, they recorded all the information voluntarily given them by salesmen.

On the second call they were instructed to be "active buyers"; that is, to ask leading questions, and to draw from the salesman all possible information or arguments he possessed concerning the relative merits of gas and electric ranges. Again, they recorded this information immediately upon leaving. In this way they drew out all the arguments for or against gas or electricity the salesman could or would produce.

Several precautions were taken to insure the validity of the results. To keep possible bias on the part of the interviewer from affecting results, they were told that the purpose of this survey was to check on how much these salesmen had benefited from the training they had been given in selling ranges. However, to be sure the point of interest of this survey would not be overlooked, they were given several suggestions as to the *type* of arguments to expect. These suggestions included:

1. Relative merits of gas versus electricity for cooking.
2. Comparative cost of operation, gas versus electricity.
3. Arguments on "modernity" of electric or gas ranges.
4. Arguments as to how long natural gas supply will last.
5. Arguments on the relative cleanliness of gas versus electricity. (Both as to cleaning the ranges themselves and smudging walls.)
6. Any additional points or comments brought out.

The interviewers were told these were only suggestions to aid them, and not to expect to hear them all or to have presentations confined only to these particular types of arguments.

As a further precaution, the work of each interviewer was tabulated separately. No significant differences were found in the work of the several interviewers.

The results of this survey showed conclusively that there was no attempt being made to question the permanence of the natural gas supply. In the 100 first calls, when the salesmen merely "said their pieces," the natural gas supply was men-

tioned only twice, both times in stores dealing in gas ranges and both times merely as a passing remark in discussing something else.

Ninety of the one hundred stores had appliances for sale, and they were called on a second time. On these second calls, the salesmen were asked about the natural gas supply. In this case, of the ninety salesmen, supposedly informed on this point and in a position to inform the general public, 15 said they knew nothing about it. Sixty-two statements were made to the effect that the natural gas supply is amply sufficient, and only five even intimated that the supply might be good for only fifteen years or less.

These facts are interesting in themselves. However, they are still more interesting in showing that even a high executive may get a fixed idea which is in many cases, as in this one, exactly contrary to the true facts about his business. And the best way to dispel such an idea is to check the facts objectively.

As might be expected from this type of interview, several by-products appeared from the tabulations. One rather significant fact about salesmen developed: men handling either gas or electric ranges exclusively averaged 8 arguments per call in favor of gas or electricity; on the other hand, salesmen handling both types offered only 4½ arguments favoring either kind. That is, salesmen handling both types offer a smaller number of arguments on either side of the question. This is probably explained by the fact that having both kinds of ranges at hand, they are more interested in selling *any* range than in pushing either particular type.

Results of survey. One result of surveys is that they whet one's curiosity. Usually they suggest further interesting and valuable problems calling for further examination. So it was in this case. Having learned a few facts about arguments being used in selling ranges, it seemed of interest to examine some of the details of actual sales. Accordingly, 31 salesmen working for a utility company were asked to answer several questions about each of the last five deals they had closed, regardless of

the price of the range or to whom it was sold. Several items of considerable interest turned up.

In the first place, it was found that 87 per cent of the prospects who bought ranges were probably interested in a range before being first contacted. Forty per cent of the prospects sold were turned in as prospects by other employees, and 30 per cent came in to the display on the floor. Only 10 per cent were "canvassed" prospects.

Second, the data showed that 71 per cent of the sales completed were closed on or before the third personal contact by the salesmen, and that 29 per cent took four or more contacts. Unfortunately nothing could be determined concerning the number of contacts made with prospects who did not buy. In view of the fact that an earlier study made in the East showed that the bulk of the business was done after the fourth contact, it would be most interesting to see if a minimum of five calls on each prospect would increase the percentage of prospects who actually bought.

Another fact bearing out the contention that more contact with the prospective buyer might increase the number of deals closed is that prospective buyers who came to the store unsolicited closed on the average with one contact less than did those received from other sources. Presumably they had partly decided to buy before coming in. It would seem to follow that in general the more remote the source of the prospect, the more contacts would be needed to close the deal.

In 51 per cent of the cases the first contact was made in the customer's home, and in 34 per cent it was made on the floor, the others being indirect. This division is, of course, to be expected, since usually we have to go out to contact prospects that we have merely been told about. In answer to the question, "Where was the deal finally closed?" we find a fairly even division. Forty-seven per cent were closed in the store, 45 per cent in the home, and 8 per cent in miscellaneous places. A recent survey of 1000 sales by a large range manufacturer shows 68 per cent, rather than 45 per cent, closed in the home.

Our figures showed 48 per cent of the deals were closed with

both husband and wife present, but only 28 per cent closed in the evening. The study of 1000 sales just mentioned shows 10 per cent more deals closed with husband and wife present, but 55 per cent instead of our 28 per cent closed in the evening. The implication here is that deals are most easily closed in the evening and when both husband and wife are present. However, again there are no available control data showing how prospects who did not buy were handled or contacted.

Having thus acquired some objective data on what salesmen were saying, and on the physical aspects of their sales, it was logical to take a further step and find out what people in general considered the particular merits either of gas or electric ranges.

Eleven surveyors interviewed a total of 776 housewives. The questionnaire forms used were revised twice after trying them out on small groups to eliminate any stereotyped answers that the questions themselves might provoke. Also, the work of each surveyor was tabulated separately to uncover any biases that might result from that source. Since the saturation of electric ranges in the area surveyed was quite low, only 2¼ per cent, 150 homes known to have electric ranges were included among the homes canvassed. This assured getting a representative sample of the opinions of both electric and gas range owners.

A rather unexpected point was brought out. Dividing the group into two parts, those preferring gas ranges and those preferring electric ranges, we find an entirely different set of comments made by each group. Those preferring gas ranges, in answer to the question, "For what reasons do you prefer gas ranges?" gave the following three main reasons:

Used to it, never used electricity	33%
Gas is faster	32%
Gas is more economical	32%

Those preferring electric ranges gave these reasons:

Cleaner	73%
Cooks better	16%
Safer	15%
As cheap or cheaper than gas	13%

And in spite of the fact that there has been a tremendous amount of publicity about electricity being the "modern way," only 7 per cent mentioned this point. This may be, not because the idea of "modernity" hasn't had its effect, but because it is a little hard to express in concrete terms.

It seems to be a common assumption that people like and enjoy new things they purchase, and have little fault to find with them. This idea was not borne out. Surprisingly enough, ownership of new ranges had little effect on decreasing the percentage of people having objections to the ranges they owned. Asked the question, "Have you found any objectionable or inconvenient features in your present range?" 22 per cent of new range owners answered "Yes," and only 24 per cent of a random sampling of people said "Yes."

Advantages of survey method. These samplings of information indicate that the survey method can be very useful for uncovering facts in a variety of situations. Almost any desired kind of information may be collected by an adequately prepared and carefully administered survey.

However, it should not be felt that such surveys have "settled" anything, or uncovered any final answers. Policies change, ideas change, and people's thinking changes. Surveys are excellent for showing present trends or levels, but they never give a complete picture or tell the whole story. Fortunately, neither business nor public opinion is static. To be most efficient, the policies of business should be at least abreast of, or better still, slightly ahead of public opinion. Too many times the policies of business diverge sharply from public sentiment, sometimes with disastrous results for business. A survey can often act as an index, showing business whether it is adequately caring for the needs of the public it serves.

Altogether, a survey has several advantages over other methods of gathering information. First, surveys may be relatively fast; the desired information can be gathered and assembled in usable form in less than two weeks. The second advantage is a corollary to this: they are current; that is, they show what is going on right now. Third, they are representative; they

show what people as a whole are thinking, rather than what one or two of us may think is going on. If well selected, a fairly small sampling can give a picture of the whole population accurate to within a few per cent. Fourth, they are reasonable in cost, considering the uses to which the information can be put. Fifth, they are objective. They show, not what *we* think or do, but what is being thought or done generally. Even the biases of interviewers can be kept from creeping in, or can be balanced out or adjusted later. Sixth, a series of surveys made from time to time can show trends that are in progress, and it has even been proved practical to project these trends as much as six months into the future with a surprising degree of accuracy. Seventh, they can be a measure of our status with our customers, that is, a measure of people's attitudes, which is more satisfactory than the usual measures based on sales volume. This can be highly significant; for instance, the Psychological Corporation's Sales Barometer showed that one brand of cigarettes was losing customers to another brand. It was predicted that the dollar volume would drop, which happened; but owing to advance orders by distributors and wholesalers, the manufacturers did not feel the difference until six months later. Finally, surveys can answer such questions as, "Why do people buy this instead of that?" and "What do they buy?" Such information can be invaluable to advertisers and sales managers in setting up their future campaigns.

Precautions needed. Although this may all sound too good to be true, it has been demonstrated many times that these points are not over-stated. However, there are several precautions that must be taken or the results will not only be worthless; they will be highly misleading.

In the first place, to secure a representative picture of a situation, we must be assured of an absolutely random sampling, which is not always as easy to achieve as it would appear. Obviously, however, if we wanted to know how many Model T Fords were running around the streets of Seattle, we would not pick a de luxe hotel garage as the place to make our survey. Our sample must be scientifically balanced.

Secondly, we must avoid leading questions, which tend to force some particular answer. The lawyer wishing to demonstrate the presence of a red barn asks the witness, "You saw the red barn, didn't you?" not, "Did you see a red barn?" A "No" answer to the question, "You saw the red barn, didn't you?" tends to imply that the witness is unobservant or has a poor memory, so he answers "Yes." Surveys should avoid prejudging anything; they are after information. The importance of careful wording is illustrated by the fact that in one test survey, the addition of one word in half of the cases resulted in a 12 per cent greater "yes" response.

Third, the interview questions should be read word for word by the interviewer, and never quoted from memory; and the interview blanks should never be left to be filled out and returned.

Fourth, the number of questions should not be too great; twenty to twenty-five is the upper limit. If possible, most of them should be capable of being answered by "yes" or "no," or by a very short word or phrase. It goes without saying, of course, that the questions should all be simple.

And fifth, the interviewers should be trained in the method of getting the interviews and recording the answers to assure standard and constant procedures.

In closing, let me repeat, surveying is a psychological method, not an answer. It never brings final results; for people's ideas, beliefs, and attitudes are never final or fixed. But properly administered and evaluated, a survey can be of great value in showing us the lay of the land. Further examples of the methods used will be found in the section on social psychology.

References

1. Jenkins, John G., "Dependability of Psychological Brand Barometers, I," *J. of Appl. Psychol.*, February, 1938, Vol. 22 (1), 1-7.
2. Jenkins, John G., "Dependability of Psychological Brand Barometers, II," *J. of Appl. Psychol.*, June, 1938, Vol. 22 (3), 252-260.

3. Link Henry C., "A New Method for Testing Advertising and A Psychological Sales Barometer," *J. of Appl. Psychol.*, February, 1934, Vol. 18 (1), 1-26.
4. Link, Henry C., "How Many Interviews Are Necessary for Results of a Certain Accuracy?" *J. of Appl. Psychol.*, February, 1937, Vol. 21 (1), 1-17.
5. Link, Henry C. and Lorge, Irving, "The Psychological Sales Barometer," *Harvard Business Review*, January, 1935.
6. Rollins, Malcolm G., "The Practical Use of Repeated Questionnaire Waves," *J. of Appl. Psychol.*, December, 1940, Vol. 24 (6), 770-772.
7. Suchman, Edward A. and McCandless, Boyd, "Who Answers Questionnaires?" *J. of Appl. Psychol.*, December, 1940, Vol. 24 (6), 758-769.
8. Wulfeck, Wallace H., "A New Application of Survey Techniques to the Study of Public Attitudes Toward Contributing to Social Agencies," *J. of Appl. Psychol.*, December, 1939, Vol. 23 (6), 723-732.
9. "The Philosophy of Customer Research," Customer Research Staff, General Motors, Detroit. (Undated, but subsequent to 1933.)

Part Nine

AVOCATIONAL PSYCHOLOGY

By

HAROLD G. SEASHORE

Springfield College

CHAPTER 30

Psychology and Art

INTRODUCTION

PUBLIC interest in art is growing. Evidence for this assertion includes the employment of recognized men as artists-in-residence in universities, the appearance of best-selling volumes of prints and old masterpieces, the enrollment of football players in college courses in art appreciation, the general approval aroused by the Federal Art Project's objectives and achievements, the increase in traveling exhibits at fairs and museums, the resurgence of interest in homecrafts and arts, the painting of better-than-average murals even in crossroads post offices, the increased use of art in camps and recreation centers, and the employment of creative activity as a therapeutic aid in mental hospitals.

The cartoonist's conception of the artist as a Bohemian, a physical wreck starving in an attic, an irresponsible member of society, is giving way to the reality. Scientific scrutiny is gradually revealing objective and useful data about the nature of talent, its development and expression. Aptitude for art is now conceived to be as measurable as aptitude for a college education. Aptitude for artistic production and appreciation is probably as widely distributed among the population as are other mental traits which have been subjected to measurement.

Educational psychologists are proposing experimentally verifiable techniques for introducing both passive and creative art experiences to children. They inquire as to whether the child's first art work should stress skills and formal rules or whether it is better to begin with expressive activity, teaching the skills and rules as their need appears. The same issue needs to be settled at the advanced levels. The futility of copy-book techniques of developing artistic understanding has been established. Psychologists have even invaded the museum to study

the appreciation habits of the patrons. Lives of great artists have been the subjects of so-called psychological biographies. Though often not very scientific, these biographies reveal significant data about artists and artistry. Advertising firms use psychological techniques to test the efficacy of commercial artistic work for bringing in the buyers.

"Pure" psychology sees in this growing interest in art an excellent opportunity to study further such traditional and persistent problems as perception, color vision, and conditioning. Since all of our mental processes operate in artistic production and in observation of art works, fruitful experimental problems abound.

In the following presentation we shall first look at a study by Lark-Horowitz on the nature of perception in which it is asked, How well can adults draw common objects and scenes from memory?

Following that there will be a description of the Meier-Seashore Art Judgment Test, acquainting the reader with the methods of test construction as applied in a special field. In a ten-year research program directed by Meier, more than a score of competent students attacked the problems of the genesis of artistic ability and the nature of this talent. At the end of the chapter there is a brief review of Meier's summary of the factors which constitute artistic talent.

1. DRAWING OF FAMILIAR FIGURES BY ADULTS [1]

The problem. Whenever there is occasion to watch an adult draw a picture of some object familiar to him or try to sketch a map or a setting that he wants to explain, one will observe that he has great difficulty in giving a graphic presentation of what he has in mind and that the result will be primitive or childish. Why is it that an adult, mature in his verbal or written expression and able to give in words a fairly accurate description, fails so considerably when drawing? Is it lack of

[1] Condensed from Lark-Horovitz, Betty, "Interlinkage of Sensory Memories in Relation to Training in Drawing," *J. Genet. Psychol.*, XLIX, 1936, 69-89.

manual ability? Is it lack of memory of the objects he has been asked to draw? Or is it that he has the kind of memory which retains the object in a form unfit for graphic representation at a later time? Answers to these questions not only have bearing on the nature of serviceable perception and memory but also would give hints on the curriculum of art education in the elementary and secondary schools.

Before reading further, the student ought to examine himself with respect to the questions asked. Let him take a sheet of paper and draw a church interior, or an object which he has experienced visually, and in other ways, thousands of times.

The experiment. Seventy-four men and ninety-six women, ranging in age up to seventy years, were asked to draw the following objects: violin; chair; duck; horse; man; woman; child; flower, the one you know or like best; automobile; town, street or country road which extends away from the place on which you imagine you are standing, with some objects alongside of the street or road. These were selected, in part, because another experimenter had obtained drawings on eight of the topics from 50,000 children.

Results. Regardless of the training, general intelligence, or profession of the person tested, the drawings produced showed a strong likeness to the drawings of children who have received practically no training and whose graphic development has not been interfered with. In the earlier study of 50,000 children the experimenter had been able to classify children's drawings into three different developmental levels, namely:

1. *Schematic or formalized presentation.* The picture is neither true to appearance nor to form and nature. However, it is true and faultless according to the child's own logical knowledge; as for example, both ends of a house are shown, a pocket on a coat shows through the sleeve of the coat (transparency), or both eyes are drawn on a profile face. The drawings seem to be the sum of the child's observation.

2. *Presentation true to appearance.* Size and proportion are correct, and details are fairly correct, but the whole figure is two-dimensional, that is, without depth and fore-shortening.

3. *Presentation true to form.* The drawing is at least an attempt to produce three-dimensional effects.

The drawings as shown in four plates in the original report are mainly on the first level. The adult's sketches also are schematic and poor schema at that. The samples in Fig. 49, of course, are only a few of the nearly 1700 adult figures obtained. However, they do give some notion of the quality of the drawings and make possible comparisons with children's productions.

The experimenter made a more detailed analysis of the character of the objects drawn. Naturally ducks are more common than violins and some subjects actually lacked close-up or frequent experiences with some of the objects. But everyone has had more than casual opportunities to observe horses and automobiles. Poor drawing was not confined to violins or to other less common objects. Surely everyone has looked down a long avenue or street; to all of us this is a daily experience. Yet few adults could get perspective into the scene; some had their stores and houses lying flat along the street.

An information sheet was used with the subjects; questions 17 and 18 from this are especially significant here.

Question 17: If you had a distinct picture in your mind of the object you were to draw, did you try to follow the shape of this object instinctively[2] with your hand just as you would when copying a picture or a real object? Forty per cent of the subjects answered this affirmatively; they were trying to "copy a mental picture."

Question 18: Did you try to remember how the object was made, of what parts it consisted, how long and how wide these parts were, what was its moving mechanism, and did you then reconstruct it accordingly? Yes, said 51 per cent of the adult drawers.

Nine per cent said they used both methods. The percentages were similar for both sexes.

[2] Instinctively probably was used by the experimenter in the popular sense, for she was not talking to persons with psychological training.

Thus, only 40 per cent claim to have retraced the object from visual memory only; 9 per cent claimed to use a combined method; whereas 51 per cent relied entirely, they said, on logically interlinked, memorial experiences.

FIG. 49. A few samples of adult drawings set along children's drawings.

16. adult
17. 10-year-old
18. adult
19. 5-year-old
20. adult
21. adult
22. 8-year-old
23. 7-year-old
24. adult
25. 9-year-old
26. adult
27. adult
28. 8-year-old
29. adult
30. 7-year-old

(Figure 3 in original.)

Interpretation. Lack of manual ability does not seem to be an adequate explanation of these inadequate adult drawings. Adults are trained in co-ordination, motility, and agility. All can do the fine movements required in writing. Other studies and common observation prove that adults are able to *copy* these objects better than children can.

The fundamental problem is not merely the inability to produce *artistic* pictures but the inability to produce essentially *correct* pictures. These factors probably operate:

1. *Blurred mental pictures.* The person with a blurred image usually is unaware of the lack of precision. Since he can recognize the real object intellectually, he is sure he can reproduce it. In answer to questioning, the subjects in this experiment rated themselves high in deftness of drawing.

2. *Clear but fragmentary picture imagery or memory.* Correct details are incorrectly assembled because of gaps in memory or imagery. The false filling in of these gaps further confuses the correct details.

3. *Lack of real knowledge of the object.* This might be true, for instance, of the violin since not many persons have ever held a violin or examined it closely.

4. *Interlinkage of sense experiences.* One's memory of an object does not depend alone on form or color. It is "sense-mixed." A memory of a chair, or of a country scene, or a person is a mixture of visual, auditory, tactual, olfactory, and kinesthetic and other experiences. These sense experiences are interlinked to form a unified whole. When one tries to make a visual reproduction, that is, a drawing, of such a total experience, a sorting out of the initial perceptual experiences is not possible. The incorrectness of the drawing seems to be due to the inability of the drawer to disentangle the interlinked sense impressions and to pick out and connect those which *can be* translated into lines and colors on paper. The schematic nature of the adult drawings suggests a return to childhood memories which seem to stand out clearly.

5. *Habits of thinking in abstract concepts.* Many churches have been in one's experiences. Church is a verbal concept;

church also *means* sounds, odors, form and color, people, emotional activity, kinesthesis, other memories of complex experiences. Memory of the sheer outline of a church is a minimum quantity in the totality. Efficiency in everyday thinking demands abstractions, generalizations, often to the detriment of details. The process of learning what a church building is does not usually involve any experiences in reproducing its form and color with pencil and paper.

6. *Education stresses verbal and conceptual thinking.* With advancing education, verbal and conceptual thinking diminishes the wish, and arrests the ability, to present objects graphically. Speaking, arithmetic, reading, and oral language are considered more important in the curriculum than graphic representation. What graphic representation there is often degenerates into copy work on the one hand or such stress on purely creative work on the other as to minimize the training of accurate observation. Effective observation which will lead to effective visual memory is neglected. Having stopped early in what little experiences of graphical representation they have and having substituted a verbal technique of expression, the graphic thinking of adults has not ordinarily proceeded beyond the level of small children.

Discussion. The experiment and interpretations shed light on several topics the reader has studied in general psychology. They relate to the psychology of sensation and perception, to memory and imagery, to thinking and abstraction, to the development of attitudes and interests. The unity, the wholeness, of mental life is also to be seen.

If we assume that graphic expression is to be considered a legitimate and useful part of general education, a very practical problem arises. How can psychology aid the art educator in developing a curriculum which will result not only in growth in eye-hand co-ordination and in a desire for creative expression, but also in lifelong habits of accurate observation leading to accurate reproduction? Another question is posed: Is it possible that the drawings which are "required" in various science courses (animals in biology, flowers in botany, experimental

set-ups in chemistry, etc.) can be drawn not so much as routine copy-work but as the reproduction from memory of essential parts in proper proportion and detail? The drawings might be less geometrically accurate but there is reason to believe that if accurate but non-copied drawings were required of students, the necessary motivation to high-grade observation would result. Perhaps it is not important that sixty-year-old chauffeurs, plumbers, and dentists are like ten-year-old children in drawing ability. But this lack does seem important if this demonstrated inability mirrors a widespread lack of accurate perceptual experience and confused memories for the contents of initial experiences. Without disregarding what we know about the transfer of training, it would seem pertinent to recommend that somewhere in the educational experience of every person more stress should be laid upon accurate observation. From experiments with errors of verbal testimony in courts, we have reason to believe that inaccuracies of verbal description are of the same order as these inaccuracies of graphic reproduction, because they spring from the same feature of human behavior, namely, faulty observation.

Finally, these data should be interpreted further in connection with the résumé of the Meier-Seashore Art Judgment Test in this chapter. Especially one could discuss these findings in terms of the assertion that manual dexterity is not the *sine qua non* of artistic ability, but that the ability to be self-critical through observation of what one sees or does is fundamental.

2. THE ART JUDGMENT TEST [3]

Measuring artistic talent scientifically. Artistic talent is recognized as a complex of many abilities, attitudes, and capacities. It is so complex in its expression that one is skeptical of an armchair analysis of its essential elements. It is obvious

[3] From N. C. Meier and C. E. Seashore, *The Meier-Seashore Art Judgment Test, Examiner's Manual,* Bur. of Educ. Res. and Serv., Iowa City, Iowa, 1930, pp. 24. Numerous other papers by N. C. Meier, chiefly, "A Measure of Art Talent," *Psychol. Monog.,* XXXIX, 1928, pp. 184-199, and "Factors in Artistic Aptitude," *Psychol. Monog.,* 51, 1939, pp. 140-158.

that we cannot measure art judgment *in toto* with any simple testing instruments. We must begin, as Meier and Seashore, McAdory,[4] and Lewerenz[5] have done, by isolating some agreed-upon significant aspect of artistic talent. Meier and Seashore, whose work is reviewed in this chapter, proposed to build a test to measure art judgment, an ability which they and others consider basic to success in art.

What is art (esthetic) judgment? Meier says:

> Simply defined, art judgment is the ability to recognize esthetic quality residing in any relationship of elements within an organization. It is vital to the artist in that good esthetic judgment permits him to know when his composition is good or unsatisfactory and what might be done to improve it. It is also the basis for art criticism and underlies the appreciative aspect of the esthetic response.
>
> In the interests of clarification it should be understood that esthetic judgment is not the application of a series of rules but is something which the individual acquires on the basis possibly of some innate neuro-physical constitution.[6]

By holding that art judgment is primary one does not need to neglect the importance of good color sense, imagination, skills, or any of the score of variables which determine artistic output. Rather the stress is that *compositional* factors are indispensable in all arts. It is well known that great masters, with few exceptions, followed the practice of making trial compositions. These plans, layouts, and sketches more than anything else distinguish the master. Thus, Meier and Seashore felt, a test of fundamental artistic talent should be so constructed that we would test the ability of persons to be critical of compositional elements in art works.

Selecting a method for measuring art judgment. This ability to be self-critical of one's work, to be cognizant of the

[4] McAdory, M., *The Construction and Validation of an Art Test,* Teach. Coll., Columbia Univ., 1939, p. 383.

[5] Lewerenz, A. S., *Test in Fundamental Abilities of Visual Art,* So. Calif. Book Depository, Los Angeles, 1927.

[6] Meier, N. C., "Factors in Artistic Aptitude," *Psychol. Monog.,* 51, 1939, p. 155.

essential compositional qualities of balance, unity, rhythm and other art variables, could be measured in several ways. First, one could have the person being tested produce a series of sketches which could be judged by experts. This procedure would be time-consuming, would penalize the subject with good art judgment but poorly trained drawing skill, and would be rather subjective.

Second, one could present the subject with a collection of "good" and "poor" pictures, as previously selected by a few present-day experts. Arranged in random order, these pictures would then be sorted by the subjects into two piles. One might also ask for a verbalized criticism of each picture placed in the "poor" pile. Graded items could be developed.

Third, one might modify the second procedure still further by using the "controlled alteration" plan, which was the one finally adopted by Meier and Seashore. This method has the advantages of being objective, economical, and non-tiring. Samples of accepted pictures (works of old masters and contemporary artists, plus some Japanese prints) were redrawn in ink-outline sketches. Then, for each picture, thus accurately reproduced for essential artistic elements, a paired sketch was made which contained a distortion of some detail or section of the standard composition. Some significant compositional element was varied so as to render the picture less acceptable. Note that by this "controlled alteration" procedure one does not have to compare pictures which have different purposes, e.g., classical and modern paintings; instead one compares a time-tested "masterpiece" with a deliberate alteration of one of its fine points. This helps to avoid the difficulty of setting up "absolute standards" in art.

By proper selection of pictures for such simplification and alteration, it was possible to include in the test several examples of nearly all of the important and recognized artistic factors in composition. In the test, typical "problems" of the artist are laid out before the subject. Figures 50 and 51 illustrate two of the 125 pairs which are found in the Art Judgment

Test as finally published. Several hundreds of drawings were used in the experimental phases. Those which are included met several experimental criteria, such as, adaptability to ink-line reproduction and value for illustrating specific art principles.

Another factor for consideration was the procedure for the subjects taking the tests. Instead of just presenting the paired sketches for sorting with no guiding instructions, the subject is told of some feature in each pair of sketches to which he is to direct his attention. On the answer sheet are the letters R (right) and L (left), together with a statement of the feature to be judged. The person taking the test views each pair of pictures in the light of the instructions and makes his choice, encircling either L or R. Thus, for Fig. 50, which is number 1 in the test, the instructions are presented as follows:

L R 1. "The Rent Bill." Posture of old man.

Figure 51, which is number 49 in the test, calls attention to another matter:

L R 49. Character and arrangement of waves.

The score on the test is the number of correct choices.

Does the test measure art judgment? This is a question concerning the *validity* of the test. The authors present the following evidence.

1. Art students and faculty members in schools of art make higher scores than younger persons and persons of similar age in other fields. All art students, of course, do not necessarily have high artistic talent. There is a wide range of ability even in this apparently homogeneous group.

2. Learning—art training—cannot entirely account for these higher scores inasmuch as some eighth-grade pupils make scores above the median of the students of art. These persons are presumably those who possess latent talent. However, it is impossible separately to measure capacity (inherent potentiality) and ability (developed potentiality) nor does one need to argue that talent must be inherited and is not dependent upon opportunity.

3. The test *items* are in and of themselves valid. That is, the experimental method of selecting the pictures and the use of

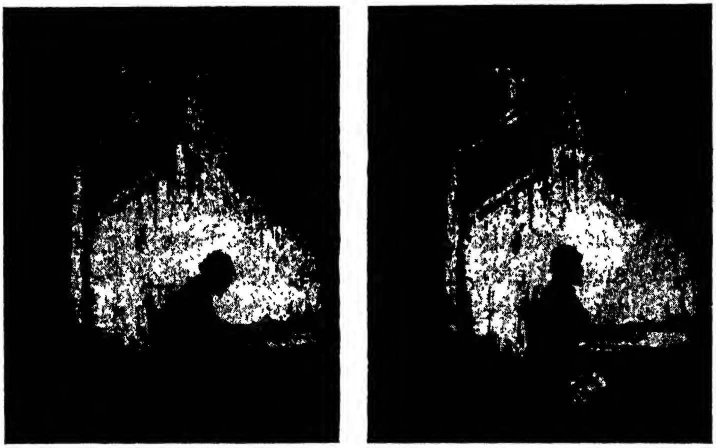


FIG. 50.

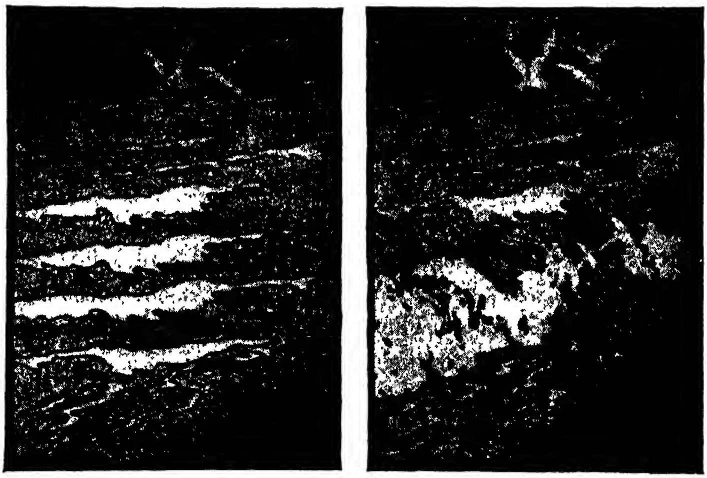


FIG. 51. An "item" from the Art Judgment Test.

experts in judging the quality of the reproductions did not alone determine the basic goodness of the art work in the pictures. The technical devices and expert judgments merely aided in selection of the better test materials out of the wide range

of possible time-verified examples of artistic quality. An "operational" rather than an "absolute" definition of goodness of artistic works is employed. Good art is what accepted critics, artists themselves, and reasonably able laymen consider to be good art. The rightness and wrongness of the pictures are objectified in terms of their acceptance in the history of art. The test has genuine validity in the sense that those who can recognize which pictures possess the time-tested qualities are those persons who have greater art judgment. The test constructors necessarily had to omit much good modern (in the strict sense) art because such works have not yet been validated in the way just described.

4. The test is not an intelligence test. Low coefficients of correlation (—.14 to .28) are found at various ages between scores on intelligence tests and scores on the Art Judgment Test. Finally a group of college teachers classifiable as superior professional adults scored much lower than another group of superior professional adults who also were artists.

Reliability. The coefficient of reliability of the test is about .75. While not as high as demanded of pencil-and-paper intelligence tests, this reliability is adequate providing the test is used for its intended purpose, namely, to discover probable talent in the general population, and is not used in an attempt to grade with sharp distinction one person from another.

Norms. Norms, in percentiles, are presented for grades 7-8, 9-10, and 11-12. Adults are scored on the 11-12 grade norms. Meier reports that the normative group of high school students is not strictly random since it included an excessive proportion of high school students in art classes; but these art students are not necessarily all high in art judgment (see Fig. 52).

Discussion. The authors urge that the test is valid and adequately reliable if used as a "dragnet" in school and other populations to discover persons who have potential artistic talent but who might be unaware of it themselves or be unknown to school teachers. Unfortunately there is no record of a sustained program of discovering art talent in a community which

would enable us to determine the practical validity of the test. That is, we have as yet no concrete evidence on a large scale that the test can keep students out of art classes who will not profit from specialized art training, who will not be able to be critical of their own work and the work of others. Also, we have no large-scale proof that it will enable teachers to select children for art instruction who otherwise might go unnoticed.

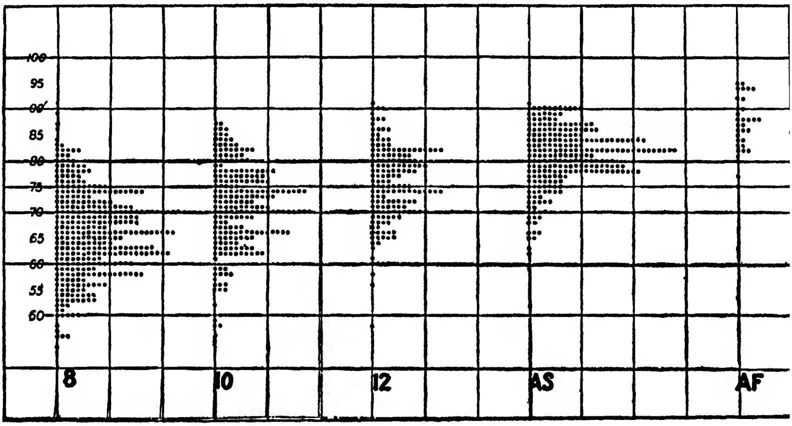


FIG. 52. Distribution of 1081 scores (experimental form) by groups.

8 = 7-8 grades; 10 = 9-10 gr.; 12 = 11-12 gr.; A.S. = art students; A.F. = art faculty.

One of the major problems in the test construction work of psychologists in such special fields as art and music is that it is difficult to interest administrators sufficiently to set up crucial experiments in whole school systems to test the utility of such guidance procedures. The Art Judgment Test suffers along with others from the fact that guidance programs are still mainly in the form of personal conferences with problem cases and as yet do not touch the more positive aspect of searching for the abilities which are most significant in every child. Such all-school studies, incidentally, would provide data for "testing the tests." Intelligence tests have been more highly developed than special aptitude tests partly because such broadside studies have been made.

The Art Judgment Test does not measure all of the abilities that constitute artistic talent, though the authors think it gets at the most crucial of these.[7] Other investigators agree that this test is more fundamental, for example, than the mere ability to draw, that is, to handle the mechanical materials of art.

Following the development of the Art Judgment Test, Meier and his students began a large-scale research under a grant from a foundation, to answer such questions as: What is artistic aptitude? When and how does it appear in children? How is it best nurtured? What crucial variables function to promote or deter growth toward artistic competency? These experimental and case-study researches have been published in various journals. The findings summarized in a paper by Meier are condensed in the next section. Meier concluded that there are six main variables which determine artistic aptitude and that all of these in turn are determined both by heredity and environment. An artistic "genius" is one who possesses all of these six factors to a high degree.

3. FACTORS IN ARTISTIC APTITUDE [8]

The finding of greatest interest is a new theory of talent which, for the first time, clearly indicates the specific interaction of the inherited aspects of talent with the learned phases. There can be no adequate understanding of artistic aptitude *without taking both into account.* The view presented herein

[7] Since this chapter has been prepared, Meier has announced Part I of a new series of tests, to be called *The Meier Art Tests.* Part I is a revision of the Meier-Seashore test which has just been reviewed. The new test is shorter, since certain non-discriminating items have been removed. Also, certain more efficient items have been weighted in the scoring. Meier announces that Part II will measure Creative Imagination and Part III Esthetic Perception. These three tests will measure three of the six major factors in artistic talent which are described in the next section. The tests are published by the Bur. of Educ. Res. and Serv., Univ. of Iowa, Iowa City, Iowa.

[8] Meier, N. C., "Factors in Artistic Aptitude: Final Summary of a Ten-year Study of a Special Ability," *Psychol. Monog., 51,* 1939, pp. 140-158. The summary here given follows closely Meier's words, except for certain omissions.

places the greater emphasis upon heredity, but extreme care must be taken by the reader to note the particular aspect of heredity involved—*constitutional stock* inheritance—not direct inheritance, in the commonly assumed sense, from parents.

There is hence presented for the first time a theory (for which there is considerable substantiation) which suggests a unique and peculiar *interlinkage* of factors that exhibit both inherited and acquired characters.

Artistic aptitude is viewed as resting upon the possession of six factors: manual skill or craftsman ability, energy output and perseveration in its discharge, general and esthetic intelligence, perceptual facility, creative imagination, and esthetic judgment. It is readily recognized that these are not mutually exclusive categories but are general terms descriptive of a number of recognizable functions which overlap considerably and are not strictly independent variables.

Manual skill. Obviously no work of art is possible without some manual skill. This ability is regarded as primarily inherited, *but not inherited as a skill* directly from parents. It comes as a phase of general constitutional stock inheritance from a *line* of ancestry which in the individual members may not necessarily include artists but *does include a comparatively* large or above average number of individuals having *craftsman ability*. The line, traced back through a half-dozen or more generations, may have included such occupational interests as toy making; wood-carving and cabinet making; watch and instrument making; diamond-cutting; textile design and manipulation; jewelry making, repairing or adjusting; or any of the arts; lithography, engraving, drafting and related activities.

The heredity involved is partly that of social selection in a consistent direction. Morphological characters are also involved: the individuals come from stock that is well adapted for work requiring fine eye-hand co-ordinations; they transmit the same characteristics, in some instances reinforced. Each new individual does not come into life with preformed skills;

he merely inherits the kind of neuro-physical constitution that is *readily adaptable* to the *acquisition* of such skills.[9]

Survey of talent. A small group of children identified at the ages of 4 and 5 years as exhibiting an interest in drawing, modeling or arranging has now been followed for a period of ten years; likewise a group approximately equal in chronological and mental age who had not at the same initial period shown any interest. . . . It is of utmost significance that the children in the first group still exhibit the same high degree of unstimulated interest and high proficiency in production as at any time throughout the period. In the case of the non-talented children the interest is still nominal or lacking and the performance average or less.

Among the established adult artists studied only a few failed to recall early artistic endeavors; others were able to lay out for the author's inspection water colors and paintings made during the age of five to eight years.

Information blanks filled out by 283 art students, chiefly in the high schools of St. Louis, disclosed more craftsman activities in recent ancestry (seldom going beyond three generations) than in a group taken at random from classes in commercial subjects and mathematics in the same high schools. Among the art students a total of 358 blood relatives listed such occupations as against 120 for the non-art group; taking the occupations most closely related to art (e.g., engraver, lithographer, etc.) the numbers are respectively 152 as against 37.

It is believed that the great prevalence of craftsmen ancestry among art people would be indicated in at least two ways—first, by the proportion of subjects having no craftsmen ancestors whatever and, second, by the average number of crafts-

[9] The student should review at this juncture his earlier learnings in biology and psychology on the topic of inheritance. He should remember that while Meier is stressing the inheritance of structural (morphological) characteristics, he does not deny, and surely would not minimize, the fact that the same craftsman ancestry provides for the children a *cultural* inheritance of family traditions, interests, training, and opportunities.

men ancestors for the individuals in each group. The results of this survey are given in Table 38.

TABLE 38

CRAFTSMAN ANCESTRY OF ARTISTS, ART STUDENTS AND THE GENERAL POPULATION

	N	*N (zero)*	*Ave.N.Cr.*
General Population			
Unselected college students	153	36%	2.05
Unselected high school students	23	35%	1.61
Totals	176	35.8%	2.00
Art Population			
Artists (limited sample)	58	15%	3.59
Art students—art schools	282	9%	4.74
Art students—colleges	230	13%	3.98
Art students—hs. and n.s.	43	13%	4.07
Art staff—engraving firm	31	6%	5.64
Totals	644	11.02%	4.37

N = number of subjects.
N (zero) = number having no known craftsmen in ancestry.
Ave. N. Cr. = average number of known craftsmen in ancestry.

The case of Loran Lockhart. This case is difficult to explain in any other way than by the acceptance of the theory now presented. This boy, blind up to his seventh year, was enabled to have perfect vision by a series of operations for double complete cataract. He then began drawing and painting at a level equal to or above normal children of his age. This ability has continued. A thorough investigation failed to reveal any possibility of environmental help or stimulus. Presence in the ancestry of craftsman ability and other aspects of his general constitutional inheritance provides the only apparent means of accounting for the facts of the case.

Energy output and perseveration. In summary it may be said that the artistically competent child discloses early in life a proclivity frequently to spend concentrated effort in work on art activities and does so in preference to almost any rival interest. At the early teen age the activity may temporarily

follow other interests than drawing or painting, such as model building or specimen collecting and mounting. But even these are closely related to the craftsman pattern of activity and serve to enhance the possibilities for development of creative imagination in later endeavors. The same tendency toward deep concentration and care for detail is strikingly in evidence.

In the adult artist the same characteristics prevail. From the extensive notes and observations gained by personal contacts with forty-one established artists the same ability to focalize great energy upon a theme until it is brought to completion is evident. Instances might be supplied of concentration and persistence that would serve to refute the popular stereotype of the artist as a temperamental, emotionally unstable, long-haired dreamer.

Esthetic intelligence. The study of the relation of general intelligence to artistic ability disclosed a definite tie-up between the two. The relation was in evidence in the high school field but was prominently in evidence in the ascertained scores of fifty-one nationally known artists of various types and of definitely high standing.[10]

Both in the case of the talented child and the adult artist superior intelligence conditions the rate of development and the functioning of other factors. It ordinarily determines the artist's competency in handling a given theme and the adequacy of his treatment. Other things being equal, it may mark the general effectiveness of the work on the whole. It may partially mark the degree of originality (entering into creative imagination) and the range of possible ways of treating a given subject.

Perceptual facility. By this factor is meant the relative ease and effectiveness with which the individual responds to and

[10] The test which was used (Otis Self-Administering) does not permit further analysis of the artists in terms of Thurstone's Primary Mental Abilities. Meier, presumably on other evidence, states, however, ". . . it is possible that the more detailed analysis of mental functions now being studied may show the artist type of mental habits probably stronger in visualizing, speed in perceiving, and possibly several more, than in others like facility with numbers or verbal fluency."

assimilates experience which has potential significance for present or future development in a work of art. The talented children not only "carried away" more of identical objective material visually experienced but also retained it in approximately the same ratio. The artistically superior child is thus one who "drinks in" more of a vacation trip, movie, or graphically presented story and retains the impression better, than does the "average" child. His perceptions of this type are therefore more adequate, and his memory for visual experience more lasting and to a greater extent available for recall.

With the adult artist this facility takes the form of more realistic and adequate "note taking" when in the presence of a scene, interesting "character" or imagined or reconstructed historical episode.

Creative imagination. Inasmuch as this term is in wide use and is perhaps little understood by its users, it is proposed that the simple designation be given as the ability to utilize vivid sense impressions effectively in the creation (organization) of a work having some degree of esthetic character. There is no need for mystification, since one does not construct without some basis for such construction. That can come only from one's experience, or as is usually the case, from composites of experience.

An example of creative imagination, motivated by an emotional initiation of the theme itself, characterizes the production of Grant Wood's *Daughters of Revolution.* Irked by public criticism of persons whom he regarded as good Americans, Mr. Wood proposed in his own mind to construct a satire which would have the broad significance of depicting the contrast between comfortably housed, elderly ladies who discuss people and issues at teas, with the actual hardship experienced by ancestors five or six generations removed. Hence the employment of contrast of color and the inclusion of Leutze's *Washington Crossing the Delaware.* It is to be noted that the incentive for this picture was a succession of news items; the creative part wherein imagination functions is in the manner in which Mr. Wood sought out and utilized photographs which

would serve as a vehicle for the satire. The three faces in the picture are constructs though based upon a study of many photographs, the actual identity of the persons being unknown.

Esthetic judgment.[11] It is not to be assumed that artistic aptitude depends upon high ratings in all six factors. We may expect that the ultimate progress of the individual is somewhat related to and conditioned by the degree to which he possesses most if not all of the six factors.

Inasmuch as the interlinkage theory set forth touches upon the nature-nurture controversy, it is desirable that certain aspects be clearly understood. The writer [12] has long shared the conviction of careful students of this problem that the explanation of psychological phenomena is not to be found in an all-or-none explanation. It is as wrong to assume that artists are "born" as it is that artists are "made."

In the case of artistic capacity as investigated over a fifteen-year period, such evidence as is presented and checked against the known view-points of present-day biology seems to suggest that some aspects of the capacity are largely attributable to the factor of stock inheritance and others to be more attributable to learning but the writer wishes to point out that the hereditary factor referred to is not heredity in the sense of direct inheritance; furthermore, that the environmental aspect is not environmental influence in the usual sense, but a relationship between the individual and his esthetically significant environment wherein the individual himself takes the initiative. Nature and nurture are here not separate elements since neither acts directly but rather interacts in a dynamic, total situation. The six factors outlined above are therefore more a series of conditions which, when present, interact with the energies of the individuals to develop his artistic competence.

It is the thesis of the writer that the person with these six factors can bring this end about and that the person without these factors cannot bring this about to any great degree. The

[11] Discussed in detail in the preceding portion of this chapter.

[12] The "writer" here and later is Meier, whose ideas are being summarized.

position is, therefore, a deterministic point of view—deterministic in the sense that certain neuro-physical and developmental factors seem to be normally a *pre-condition* for the rest of the total development and that these pre-disposing conditions are not present equally in all persons nor if absent can they be established.[13] The nature-nurture aspect is hence anything but a simple matter. The factors are not only interlinked in the gross aspects, but they are interlinked and conditioned one with the other in a dynamic sense—*i.e.*, the interrelationships may change with time and may exist in varying potencies with different individuals. There are probably no two individuals who present identical composites of factors to begin with and these composites are probably different at each stage of development. It is believed nonetheless that in all cases the *general* pattern is that described above, involving the factors of motor skills, the volitional-temperamental traits of energy-output and perseveration, intelligence, habits of perceiving, special utilization of imagination and a special disciplining of judgment and critical processes.

[13] On this last phrase Meier might find some contrary evidence in the work of his colleagues in the Iowa Child Welfare Research Station as reported in Section V of this book. We refer to the several studies in the field of intelligence which indicate that intelligence quotients are not as static as once thought. In the Thirty-ninth Yearbook of the National Society for the Study of Education, dozens of papers argue the matter pro and con. Those who hold that intelligence does respond favorably to stimulating environments, lay special stress on the importance of opportunities for creative activity in such environments. Is this the essence of growth in artistic ability also?

Variability in Pitch in Artistic Singing [1]

AN EXPERIMENTALLY constructed psychology of musical esthetics is the long-range goal of a number of research investigators in a few laboratories. For centuries, esthetics, defined as the formulated principles of art, has been the peculiar province of the philosophical theorists on the one hand and the practical critics and performers on the other. Armchair notions, valuable as they are, have ruled the thinking. Objectivity has sometimes been strikingly absent. Cultism and mysticism have determined the language not only of the artists themselves, but also of those who have tried to explain or even adequately describe art. Art tends to be esoteric. No satisfactory body of systematic knowledge of what constitutes artistry in any of the art fields has yet emerged. There have been non-experimental estheticians of great insight and doubtless there is much validity in the extensive folklore of beauty.

Before a truly valid science of esthetics can develop, with positively stated principles which are verifiable by other investigators according to the best methods of modern rigorous thinking, a large amount of preliminary spade-work will have to be done. Some has been done. Here and there isolated researches and in a few laboratories sustained programs of investigation have pointed the way which this psychological contribution to the field of art will take. One of these groups has been working in the Iowa laboratories. Some of the early work was overshadowed by the emphasis upon the measurement of

[1] Condensed from Harold G. Seashore, "An Objective Analysis of Artistic Singing," *University of Iowa Studies in Psychology of Music,* IV, 1937, 12-157, and from Harold G. Seashore, "Variability of Pitch in Artistic Singing," *Proceedings of the Music Teachers' National Association,* 1938, 66-80.

musical talent, but in the 1920's and 1930's particular stress was placed upon analytical studies of musical performances, ranging all the way from studies of small children and primitives to concert performers of international merit.[2] One of the larger of these studies was the five-year investigation of Harold Seashore, one part of which is the subject of this chapter.

To suggest the breadth of the field of research in musical esthetics, the following job analysis is presented:

1. Extensive objective descriptions of musical performances in all kinds of cultures. Even if we choose to make our Western culture the artistic standard, it is necessary for us to engage in anthropological and genetic research in many cultures and in all ages and levels of society.

2. Extensive objective descriptions of the reactions evoked in listeners in all kinds of musical situations. Investigations of widespread emotional reactions, more subtle feeling-tones, and the intellectual activity of the hearer are involved.

3. Experimental modifications of behavior, both in musical performers and in listeners, must be attempted to develop the laws of perception, growth, and learning which apply to musical experiences.

4. Correlated investigation into the nature and development of talent are clearly indicated.

5. A whole field of applied psychology will then be opened, involving scientifically applied principles of selecting and guiding persons for various kinds of musical training, both as performers and as consumers; application of the best principles of pedagogy; and the use of music for social control, as in morale, propaganda, advertising and the like.

In the present review only analytical descriptions of singing performances can be reported. Further, only the sections dealing with variability of pitch will be presented. Parallel studies for other musical instruments and for other aspects of a performance, such as rhythm, phrasing, time, intensity, and so on, are reported in the monograph from which this typical excerpt

[2] Much of the research has been summarized by Carl E. Seashore in his *Psychology of Music*, McGraw-Hill, 1939, in which the original researches are adequately cited.

is taken and in other publications of the same laboratory.[8] Since few psychology textbooks provide adequate background for this review, a few introductory explanations are in order. The broad principles are discussed in relation to pitch performances of singers.

Artists strive for perfection. Audiences, critics, teachers, and coaches alike have demanded an approach toward perfection in the artist. Definitions are often hazy, and certainly the critics disagree as to the attainment of perfection. It must be granted readily that, by and large, trained critical listeners are the best judges of quality of singing that we have today. But their criticisms are based mainly upon what they like. It must also be assumed that ultimately the professionally competent critics, including skilled performers and the trained audience, are right; what is liked is right esthetically. But between expressing what one likes and describing that performance accurately there is a great gap marked by dozens of illusions, biased mind-sets, and frailties of human perception.

Variability: A fundamental art principle. In the singing of a song there are so many musical factors which are continuously interacting that the attainment of any sort of perfection in any one is hardly to be expected. Yet the absolutist in musical criticism speaks in terms of absolutely correct pitch and true intonation. One of the first things the scientific investigator of music has learned is that in no performance is the singer on true or exact pitch more than momentarily. In fact, singers are unbelievably "off-pitch" in the performance of even simple songs. *Variability is the rule.*

1. The conventional musical score—the composer's documentation of the tonal sequences which he feels will express beauty, emotion, and meaning—is for the singer only a schematic reference about which he weaves, through continuous variations in pitch, a

[8] Seashore, Carl E., *ibid.;* Wilbur S. Schramm, *Approaches to a Science of English Verse,* University of Iowa, 1935; and various articles from the Phonetic Laboratory published in the *Quarterly J. of Speech* and the *Archives of Speech.*

nicely integrated melodic unity. In a very real sense, a singer never sings on pitch.

2. With respect to precision and rigidity, singing falls at a mid-point between two extremes: Players of rigidly pitched instruments, such as the organ or piano, can render the composer's pattern of pitch exactly and repeatedly, whereas readers of verse and prose, in which the melodic pattern is more indefinite, express themselves with the greatest variability and non-repeatability.

3. These artistic deviations from the compositional statement of melody are of three main forms: (a) the pitch vibrato appears as a periodic variation in frequency of sound waves in every tone; (b) the general tonal level in performance varies above and below the correct pitch level and affects such musical factors as interval, melodic direction, and legato flow of tone; (c) the transitions between tones show several forms of variation of pitch in attacks and releases and in portamenti.[4]

Students of esthetics have discussed this problem of variability in relation to all of the arts for years. The esthetic principle may be stated: *artistry consists in deviations from the regular, the perfect, the rigidly exact.*[5]

Paradoxically, artistic perfection arises out of technical imperfection. Carl E. Seashore, who has been a leading research investigator in these areas, and one of his students have expressed the art principle with reference to singing in these words:

. . . The principle involved is a well-recognized theory of art. The new feature is the objectifying of it in terms of quantitative

[4] Seashore, Harold, *op. cit.*

[5] This principle underlies the generally greater acceptance in the graphic arts of pictures which are not photographic in detail, in sculpturing of bodies which are not perfectly anatomical, and in music of performances which are not mechanical. Even on the mechanical piano the listener is permitted to vary the tempo and dynamics of the performances originally edited by the artist. Most of us have scant tolerance for the bleating tremolo stop of the organ. As yet, no mechanical violin has been accepted outside of third-rate ice cream parlors. Doubtless, the general preference among listeners for musical performances on the strings and by the human voice rests upon their greater flexibility whereby artistic deviations from perfect performances are made possible.

measurements. In music and speech, pure tone, true pitch, exact intonation, perfect harmony, rigid rhythm, even touch, precise time, play a relatively small role. They are mainly points of orientation for art and nature. The unlimited resources for vocal and instrumental art lie in artistic deviation from the pure, the true, the exact, the perfect, the rigid, the even, and the precise. This deviation from the exact is, on the whole, the medium for the creation of the beautiful—for the conveying of emotion. That is the secret of the plasticity of art.[6]

Singer-Song-Listener. In the singing situation there are three major units to be investigated: the Singer, the Song, the Listener. The relationships can be shown in a tabular form which, though not entirely adequate, does aid us in clarifying our analytical studies of singing.

FACTORS IN A SINGING PERFORMANCE

The Singer	*The Song*	*The Listener*
Production	*Performance*	*Reception*
(Psychophysiology)	(Sound Waves)	(Psychophysiology)
Vocal apparatus	Pitch (frequency)	Auditory system
Neuro-physiology	Loudness (intensity)	Neuro-physiology
Sensitivity	Timbre (wave form)	Sensitivity
Motor skills	Duration (time)	Motor response
Interpretation	*Interpretation*	*Interpretation*
(Psychology)	(Analysis)	(Psychology)
Perception	Norms	Perception
Cognition	Individual differences	Cognition
Emotion	Variability	Emotion
Action	Laws of artistry	Action
	Musical form	

The song is the bridge between the performer and the listener. *The song is the basic musical factor; it is the art object. All that is musical is conveyed from singer to listener on sound waves.* For scientific purposes the song is physical; it is a sequence of measurable sound waves.

Much of the confusion in the development of techniques of singing and of standards of judging singing has arisen because

[6] Seashore, Carl E., and Metfessel, Milton, "Deviations from the Regular as an Art Principle," *Proceedings of the National Academy of Sciences,* Vol. 2, 1925, 538-542.

the facts which apply to one of these three units of the total situation do not apply to the others. One case of this confusion is apparent in the classical argument concerning the emotional status of the composers and performers and the emotional responses of the listeners. Certainly there need be no expectation of more than a gross similarity between the origin and the end of the musical sequence. In fact, there are times when the listener may be having emotional reactions quite out of keeping with what the performer intended. There is also no simple relationship between the physical factors in the objective song and the auditory experiences which it evokes. There is no simple one-to-one relationship between the physical and the mental experiences. A most obvious musical example of this is that, while we may know that sound is composed of vibrations of the air-medium, we do not hear vibration but rather tones of certain pitches. Likewise, only on rare occasions do we hear overtones as such; typically we hear the tenor quality, a violin string, the clarinet, the vowel *ah,* and so on.

It is a fundamental tenet of these scientific approaches to singing that whenever a statement is made there must be a clear understanding as to what part of the total situation called singing it applies. Seashore's factual analysis is from the physical sound waves, *the singer's art products*. At times he attempted to show how these objective data are related to the experiences of hearing singing, but the final story of such relationships is not yet written.

Singers' performances are "dissected" in the laboratory. For the studies upon which this research is based, Seashore used such singers as Tibbett, Crooks, Marsh, Baker, Homer, and others. Their performances on Victor Red Seal recordings were used. Also he studied the performances of well-accepted local singers in a university community. Professor Arthur Kraft of Chicago sang for him in the laboratory-studio. Only concert songs in legato style were analyzed. From the performances of these singers, who represent a good sampling of contemporary American concert singing, he obtained "phono-

photograms" of the sound waves which comprise their artistic productions.

A phonophotogram is an objective photographic recording obtained by modern electrical and acoustical devices which permits detailed and complete measurement of the variables in the sound waves. There are four and only four variables in the sound waves: the *frequency* of vibration gives rise to the experience of *pitch;* the *intensity* of sound determines largely the experience of *loudness;* the *form* of the sound waves is the primary physical basis for *tone quality;* and the *duration* of sound determines the temporal organizations of *time* and *tempo*. The measurement of these variables and an analysis of their interrelation give a complete picture of what the singer actually does in his performance.

Truly complete analyses have not been made as yet, but sufficiently elaborate studies are available to permit a discussion with objective surety of the variability of pitch in fine concert singing.[7]

The "musical pattern score." The standard printed score which is used to report what notes and words are to be sung and the relations between them is not usable for reporting the exact performance of the singer as phonophotographed. The Iowa group over a period of years developed a "musical pattern score" which enabled them to report in detail what occurred during each fractionated second of the actual performance of the singer.[8]

[7] It should also be stated that singers and critics must not be held accountable for not having made earlier the discoveries to which this research contributes. They have had to await the growth of modern engineering science and methods of electrical transmission and recording as well as a growing interest on the part of the scientists themselves. Today physicists, phoneticians, acousticians, physiologists, anatomists, and psychologists have adequate techniques for providing a scientific basis for modern musical criticism and teaching.

[8] Separate recording and reporting devices are necessary for analyses of tonal quality. The illustration to be discussed, Fig. 53, also does not show the graph of the intensity pattern of the singer which is ordinarily included when the singer performs unaccompanied. In this case the or-

Figure 53 is a "musical pattern score" of Lawrence Tibbett's singing of "Drink to me only with thine eyes," Victor Recording, 1938. This performance was phonophotographed in the laboratory under good acoustic conditions for reproduction. The graph presented in Fig. 53 permits detailed study of several significant musical variables in Mr. Tibbett's production.

1. Melody is a matter of pitch. Pitch is graphed on a scale of equally spaced semi-tones. Since pitch is being measured in tonal distances, an equal distance for each musical semi-tone is required.[9] In the figure, notes are placed on the score; the reader can soon recognize the melody and sing it through on the new pattern. The "wiggly lines" at the beginning of each note represent the *actual pitch performance* of Tibbett plotted with a fineness of one-tenth of a musical tone (one-fifth of a semi-tone). To become acquainted with the graph, consider the first tone. Tibbett was required to sing an eighth note on G. He began on E below G, raised his pitch to G, and hovered around G with five vibrato cycles (defined later). The graphed line gives a much more analytical story of just what Tibbett did than the mere statement that he sang *"Drink"* on the opening note, G.

2. Time. The little black notes are placed in the "musical pattern score" at the place *in time* when the given note was begun by the singer. Technically, all eighth notes should be equally spaced, and similarly other notes should be proportioned in time. The *actual time* performance of the singer is here plotted. The vertical black lines indicate seconds; ten seconds are plotted on one strip (staff) of the graph. Within

chestral accompaniment could not be dissociated from the singer's intensity pattern. (In the case of pitch, such dissociation can be made if the accompaniment is not too loud.)

[9] In ordinary musical scores the spacing is irregular, involving sharp and flat signs to indicate half distances which are not actually printed. Sometimes a distance from a line to a space is a whole tone, sometimes a semi-tone. This cultural hand-me-down seems to work fairly well in ordinary musical learning and performance. It certainly is illogical and probably could be revised to facilitate better learning, but this is a problem for the educational psychologists.

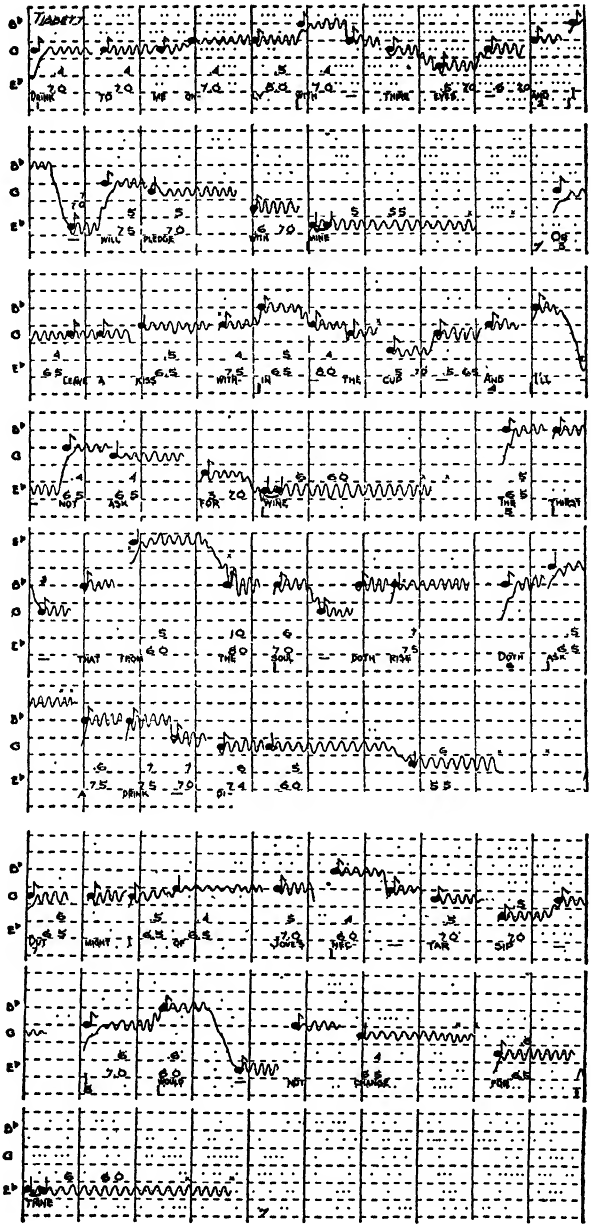


FIG. 53. A phonophotograph of the singing of "Drink to Me Only With Thine Eyes" by Lawrence Tibbett. See text for details of reading the graph. (From H. Seashore, *op. cit.*, pp. 29-30.)

each second, the small dots mark off one-tenth second units. Consider the first tone again. It lasted 1.15 seconds, followed by a pause of .15 second. The third note (on the word *me*) is also an eighth note, but it lasted only .6 of a second, less than half as long as the first note which was of equal value in the published standard musical score.

3. Phrases are numbered at the bottom of each staff.

4. Measures are indicated by short heavy vertical marks along the bottom of each staff. Again, measures are theoretically equal, but a little study indicates that in performance they are not of equal duration.

5. Syllables are printed just below the notes, thus helping to indicate the spacing of the syllables in time.

6. The small numbers just above the words are the average extent of vibrato (upper number) and average rate of vibrato (lower number) for each tone.

This musical pattern score from Tibbett is just one of hundreds of similar representations of actual musical performances which have been documented in the laboratory over several years from all sorts of singers, ranging from children to adults, untrained to trained adults, primitives to Metropolitan Opera stars; and, among the primitives, from the lesser singers to those who rate high. The same technique is used in studying speech performance, from straightforward discourse to the emotional dramatic readings of fine modern actors.

All artists sing with vibrato. The most obvious feature of a singing performance is the vibrato. In Fig. 53, vibrato appears as the "wiggles" [10] in the pitch line. Carl Seashore's empirical definition of this basic phenomenon is now accepted. It replaces the hundreds of mystical and non-observational definitions of the nature of vibrato which have abounded in musical literature since the middle ages.

A good vibrato is a pulsation of pitch, usually accompanied by synchronous pulsations of loudness and timbre, of such extent and

[10] These "wiggles" are not the sound waves, but plottings of sound waves as converted into musical notation and intervals.

rate as to give a pleasing flexibility, tenderness, and richness to the tone.[11]

Regardless of the controversial and subjective literature condemning, condoning, and praising the esthetic value of the vibrato and its role in the production of good tone, the evidence now is certain that *all contemporary artistic singers sing with vibrato all of the time in all songs and on every tone.* Most adolescent children who sing freely and with abandon show it in their voices. Untrained American Negroes and Indians exhibit it as a natural part of their vocalization when they sing according to the best standards of their people. In concert and operatic singing the vibrato occurs continuously in every tone, both in the sustained portions and in the transitional attacks, releases, and portamenti. Wherever there is artistic adult vocalization, there is vibrato.

The variation in pitch in the vibrato is considerably greater than even the most radical judgment of a subjective listener would suggest. In fact, rarely does the listener hear pitch variation at all. The average singer's fluctuation of pitch is approximately a musical semi-tone. These periodic variations in pitch occur at an average rate of six to seven cycles per second.

A large body of subsidiary findings concerning the vibrato also can be summarized. Vibrato cycles vary among themselves. There is variation in rate and extent from one vibrato cycle to the next, from the cycles in one tone to the next, from one song to the next, and from one singer to the next. While a singer's average rate and extent are quite stable, the rate and extent of single tones in repetitions of the same song vary considerably. The rate and extent of vibrato are not very significantly related to pitch level, loudness, or length of the tones, except that there is some tendency for longer tones to have narrower pitch excursions. What is more sur-

[11] Seashore, Carl E., "Psychology of the Vibrato in Voice and Instrument," *University of Iowa Studies in Psychology of Music,* Vol. III, 1936. This entire volume is devoted to experimental analyses of one musical phenomenon. Historical notes are included.

prising is that the evidence shows that *no typical rates and extents are used for expressing specific emotional contents in the lyric.*

On the basis of extensive studies and upon theoretical considerations, Seashore concludes that rate and extent of vibrato are not voluntarily determined by the singer for each tone which he sings but that *vibrato is a natural physiological concomitant of vocalization when singers perform naturally and in acceptable form.*

It would be desirable to elaborate the implications of these facts by showing the relations of vibrato to tone production, theory of placement, and to many related problems. There are complicated auditory phenomena which operate to give the vibrato its musical value. These could be discussed from the psychological angle. Suffice it to say that the listener is the subject of numerous illusions of hearing by which the physical periodicities of pitch in the vibrato are translated into a pulsing quality of tone which is not only tolerable but also a distinct contribution to the richness and mellowness of tone.

In summary, the *singer's vibrato* is recognized by him as simply the feeling of producing good tone; the *physical vibrato* in the sound waves of the song is an oscillation in pitch at certain rates and of certain extents; the *listener's vibrato* is a pleasing pulsing quality of tone. There is no question but that the vibrato gives the voice greater flexibility, more fluidity in legato passages, more "pearliness" in florid runs, and more chilling power in highly intense passages.[12] The major pitch variation in artistic singing is the vibrato. Although the vibrato has traditionally been considered a musical ornament, it must hereafter be appreciated as a basic musical factor, as one of the prime essentials of good tone production.

Artistic singers flat and sharp their tones continuously. Having discussed the major form of pitch variability, we now

[12] One can observe the effects of vibratoless tones by listening to the strident, quickly fatigued, inflexible, and "off-pitch" singing which is common in the immature or forced voice. Or, compare the relative pleasantness of violin tones played with and without vibrato.

review the second form, the flatting and sharping of tones. A singer in his training is drilled to recognize the importance of true pitch, accurate intonation, and perfect intervals. Books written by experienced masters for the young vocalist emphasize this quality of musical performance. All of us recollect having squirmed in the audience while listening to some particularly bad case of flatting or inaccuracy of interval. Because we all know that singers are sometimes off-pitch, and because we know that being on-pitch is a prime requisite of all good singing, we are taken back a little when Seashore asserts that thus far *no singer has been found whose singing is not flatted or sharped a good part of the time.* This problem is one in which the *physical* factors and the *psychological* factors must be clearly delineated because the relationships are very complicated. He is not referring here to a performer being *heard* as singing off key; he is stressing the fact that in the train of sound waves set up by the singer's vocal organs, one cannot discover rigidly accurate tones and intervals. Figure 54 shows some of these deviations of one singer.

From a significant study of many singers in simple concert songs, it is known *that only about 20 per cent of the tones are sung on correct pitch throughout their duration and that about 25 per cent of the tones are never on correct pitch at all.*[13] The other tones are characterized by inflections of pitch whereby they are on-pitch part of the time and off-pitch part of the time. The average error in pitch level is about one-tenth of a whole tone; this error is much greater than would be tolerated in single tones in the studio outside of lyrical context, or in the tuning of a piano or violin. These values are based upon the measurements of the main bodies of the tones and do not include the pitch variations in the transitional elements which will be discussed later.

Of significance to the musical world is the objective description of the inflections of the pitch levels of the singer's tones.

[13] The phenomena to be discussed in this section are all illustrated in Fig. 53. In fact, Tibbett is quite typical in that he is neither extremely accurate nor extremely erratic.

A song is a sequence of tones molded into a melodic unity. Unlike the performer of such a rigidly pitched instrument as

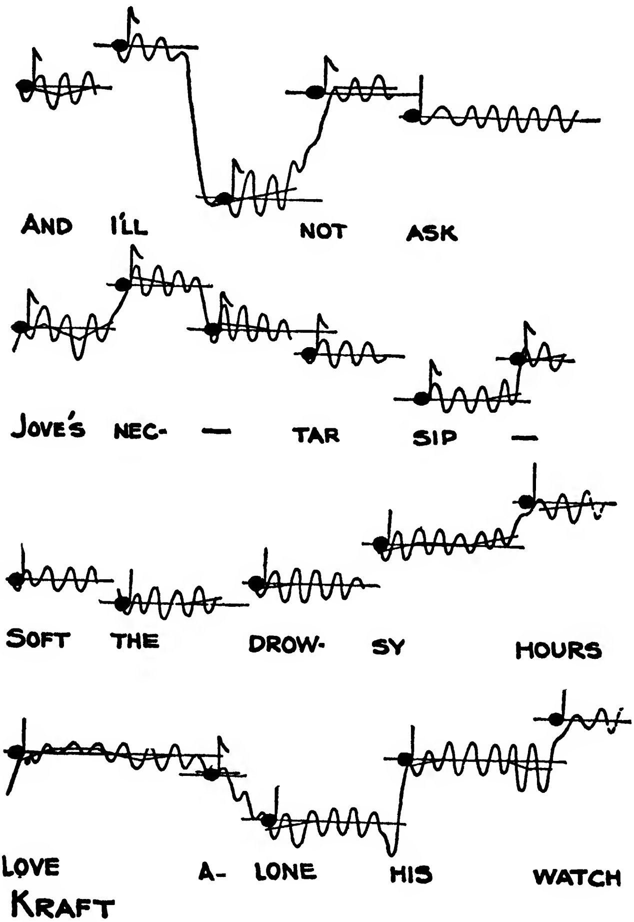


FIG. 54. Samples of tones from Kraft to illustrate deviations of mean-pitch-lines from the expected pitch. The mean-pitch-lines are defined by the lines drawn as mid-lines through the vibrato cycles. (From Harold Seashore, *op. cit.*, page 50.)

the piano, the vocalist is not confined to the "stair-cased" melody indicated by the score. The vocalist can use a "molded contour" type of tonal progression along the melodic line. Con-

trary to the edicts of writers on the subject who insist on clear-cut breaks between tones in the melody and upon discrete steps from tone to tone, the facts are that the best singers today mold their melodic flow of tone by continuous and elaborate inflections, with the result that rarely can it be said that the whole tone is sung upon correct pitch. It is recognized that these inflections of the main bodies of the tones are a basic element in legato singing.

However, it does not happen that the molding of the melodic line is as systematic as one might at first anticipate. One might suppose that if a second note were higher than the first, the singer would end the first note a trifle sharp and begin the second one a little flat. Likewise, if the melodic line were descending, one might suppose that the singer would end the first note somewhat flat in anticipation of the next tone. It has been discovered that about 30 per cent of the tones are uninflected; that is, the general pitch levels of the tones are even, but not necessarily on correct pitch. Of the remaining 70 per cent, over one half, or about 40 per cent of all the tones, are characterized by rising inflection of pitch. Falling inflection is found within a tone in less than 10 per cent of the tones. The remainder show pitch levels that are erratic; that is, their general pitch level is changed in inflection several times during the duration of the tone. Naturally such erratic patterns are found mainly in longer tones. Since the ascending and descending intervals in any large group of songs are about equal in number, it appears that pitch inflection in a tone is not closely governed by melodic inflection. Some interesting cases appear in which a singer wishing to descend, say, a musical fifth actually will drop in pitch a quarter-tone below the desired level and then raise the pitch level to the proper value. Thus, Seashore concludes that, while this great amount of general flexibility of pitch level is significant in molding the melodic line and in effecting a legato style, *the gross variability itself rather than the systematic use of the inflections is the contributing factor to melodic smoothness.*

A related problem is that of singing intervals.[14] It has been found that intervals are sung with an average error of about one-tenth of a whole tone, that no particular sizes of intervals are susceptible to greater distortion, and that the whole matter seems to be one of general instability of pitch. Of interest are the cases where a second error in the width of an interval is made to compensate for an earlier error so that the singer gets back closer to correct pitch again.[15]

Artists make transitional "errors." The third form of variability in pitch which must be understood objectively by the musician before he can build a true esthetics of singing is found in the transitions from tone to tone. Between tones the singer either makes a complete vocal break in which he ceases to phonate, or he glides from tone to tone with a continuous pitch inflection—the portamento. In these normative studies of artistic singing, Seashore finds that about 40 per cent of all tonal connections are by means of portamento glides.[16] When two or more adjacent notes are sung with the same vowel a portamento glide is indicated in the score. *Only about one-half of the portamenti which appear are indicated by the composers in the scores; the remainder are added by the singers.* Singers add portamento glides where there are changes in vowel, where there are semi-vowels to be passed over, and in a few cases where consonants have to be distorted so as to make them more vowel-like. These added portamenti obviously are a contribution of the singer to the performance. Technically, they are errors in phonation. Whether the singing would be better without them is for competent critics and musicians themselves to decide. Unquestionably, these glides

[14] "Interval" is the pitch difference between two tones. Singing D after C is an interval of one whole tone; singing G after C is an interval of three and one half whole tones, commonly called a musical fifth. A song is a sequence of tonal intervals.

[15] Anticipating the inquiry, it may be said that the phenomena discussed were found as often in the nationally known singers as in the local singers.

[16] The reader in this instance must be particularly aware that this refers to concert songs in legato style; a song with a staccato-like melody and lyric would have more out-and-out breaks in phonation.

are part and parcel of the tonal inflections discussed earlier and are a valuable element in creating the sustained flow of tone desired in legato songs.

The foregoing paragraph refers to the 40 per cent of intertonal transitions where there is a portamento. In the 60 per cent of transitions where the singers make a "clean break" in phonation, the attacks of the tones either are level, except for vibrato pattern, or are gliding. In 35 per cent of the tones, the pitch pattern at the inception of phonation is level in the sense that there is no inflection of pitch other than the upstroke or downstroke of the vibrato cycle. In the other 25 per cent of the tones, the attack is characterized by a gliding pitch pattern. Some gliding attacks are over a half-second in duration and cover a pitch span of more than half an octave. Some gliding attacks are as brief as one-twentieth of a second and are less than a quarter-tone in pitch range. The average duration of all of the gliding attacks is about one-fifth of a second and the average pitch excursion is about one whole tone. Only in rare cases does one find falling pitch glides; 97 per cent are rising.

Gliding attacks are closely associated with the initiation of phrases. The phrase openings of six singers are shown in Fig. 55. *About two-thirds of all phrases are begun with rising pitch inflection.* The use of gliding pitch is also common where the melodic line is ascending and when the approached tone is to be a long one. In three-fourths of the gliding attacks, definite vibrato cycles appear.

Combining the portamento and gliding attacks, we observe that *about 65 per cent, or two-thirds, of all tones are characterized by instable, inflected, flexible pitch patterns at their inception.* Obviously all of the gliding attacks and the portamenti which are added by the singer are technically errors, but in a normative esthetics they are not all to be classed as errors. Musician and critic will have to get together to determine which are artistic transitions and which are not.

Discussion: Errors or artistry? Probably the reader has been asking himself which of these deviations from rigidly

prescribed pitch are sheer errors and which are artistic deviations. Also, might the same deviation be an inartistic error in

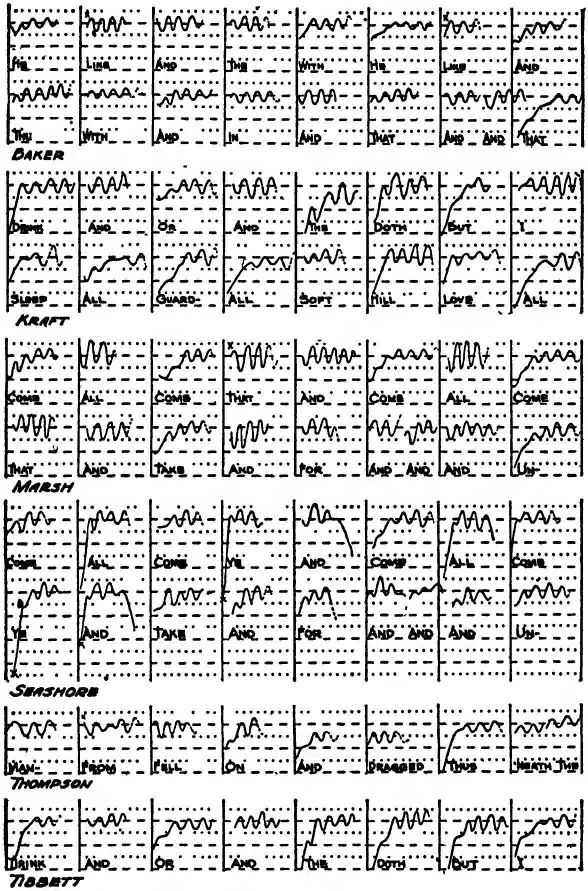


FIG. 55. The attack of phrases. The first words of the phrases in the songs of six singers are presented. The pitch patterns show many gliding attacks and a few level attacks. Some glides are very extreme. Observe that gliding attacks show the vibrato. (From Harold Seashore, *op. cit.*, page 58.)

one situation and a fine artistic nuance in another? Unfortunately, completed researches have not been carried beyond the verification of what is reported above. Just how artists and

non-artists differ in detail is not yet known with sufficient accuracy to warrant a report here. Variability is a vital component of artistry; yet variability, in and of itself, is not necessarily artistic. It is known that singers with vibrato judged to be poor or bad have wider than usual or faster or slower than usual vibrato; this suggests that extreme deviations or the lack of integration due to the combining of certain rates of vibrato with certain extents of vibrato may be inartistic. In brief, the whole story cannot yet be told.

However, on the positive side, the fact that Seashore's research, and those which led to his and emerged from it, have pointed so concretely and conclusively to the fact of pitch "inaccuracy" in the best singers indicates that much of the non-experimental literature on the nature of singing and the teaching of singing will need to be negated or revised.

When does a singer make these errors? At will? Probably he does not set out while engaged in an artistic performance to apply these errors. It is more likely that *whenever a singer performs with competence, with musical intelligence, and with emotional freedom, he will make the sort of "errors" which we have been describing*. Artistry is not technical *accuracy* but rather technical *proficiency* coupled with musicianship and emotional expression. Artistry demands deviations from the regular, the rigid, the exact. After all, a musical score is a very crude way for a composer to convey all of the ideas he wishes to convey. The singer must of necessity regard the score as the barest reference pattern.

A few speculations are suggestive of what might obtain if similar studies were made of more colorful singing such as that of the coloratura and the operatic singer, or that called for in the more emotional concert songs. Two comments seem pertinent. First, exploratory studies, mainly on the vibrato, suggest that in other kinds of singing, such as operatic arias and more staccato concert passages, the same *qualitative* varieties of pitch variations appear, though they may differ in the percentage of the time they are present or in the amount of the deviation. Second, it appears that the more the singer

expresses himself with freedom from restraint and the more he deliberately avoids being aware of technical excellence, once that technical excellence is habitual, the more radical will be his pitch variability. Thus one could expect that in the florid runs of a coloratura the portamento glides would tend to mold a melodic contour in which the separate musical steps on the score would be minimized and more of the total tonal duration would be spent on transitional portions than on the so-called main body of the tones. In short, the singer would spend more time approaching and leaving the tones than actually singing the main portions thereof. At times, in such elaborate passages one might expect the vibrato pattern itself to carry a part of the brilliance of effect. Another instance of the greater flexibility of patterns in certain types of songs is apparent in operatic and other selections where the dramatic element makes the performance almost like the reading of verse. We know from studies of reading of poetry and dramatic lines that speech approaches the quality of singing in many cases; likewise there are cases where artistic singing approaches the patterns of pitch in excellent speech.

In conclusion, this research by Seashore, as well as the others of which this is but a typical instance, demonstrates a tendency in psychology which the student should recognize as being extremely significant. Sometimes psychologists are asked why they, of all people, do certain studies. Isn't this physics? Isn't this physiology? Isn't this music? Isn't this philosophy? The answer, stated rather dogmatically, is that there are few real problems today which can be solved by the techniques and professional workers in one field alone. The psychologist who works in experimental esthetics must pursue his study jointly with the engineer, the physicist, the acoustician, the phonetician, the anatomist, the physiologist, the musician, the philosopher, the statistician, and the anthropologist. The day of isolationism in research is gone; today crucial inquiries are made by men with training in several sciences who know how to cooperate with competent colleagues in closely related, or even quite remote, fields.

The Mood-Character of Musical Phrases [1]

THE SYMPHONY orchestra is tuned, the maestro strides to his podium, his hands are raised, the opening strains of Beethoven's *Pastoral* symphony sound through the hall. To some of the thousand or more listeners it is an old musical friend, to others it is new. The present performance is superb, as attested by the attentive listening and later by the loud burst of applause and the press accolades the next morning.

The psychologist can ask a large number of questions of musico-scientific interest about such a performance. There are questions about musical talent, about composers and their compositions, about social interaction in concert halls, about emotional responsiveness, and various problems of joint interest to the psychologist, the physiologist, the physicist, and the musician. Suppose we ask just one question: What did the listeners think about during the playing? This problem of the power of music to arouse intellectual responses is an old one. If you had recently seen the Stokowski-Disney animatograph *Fantasia,* it is quite likely that your mental activity during the *Pastoral* would embrace centaurs, Bacchus, cupids, and Mt. Olympus. For others the musical phrases would evoke more natural pastoral scenes. In some it might arouse memories of specific experiences in connection with the symphony, such as a proposal on a Bermuda-bound boat while the ship's orchestra played this music. To a Kansas co-ed it might "mean" all her experiences in the open country, widesweeping vistas and rolling hills. A Dartmouth football player might recall the

[1] Based upon two researches of R. H. Gundlach, "A Quantitative Analysis of Indian Music," *Amer. J. of Psychol., 44,* 1932, 133-145; and "Factors Determining the Characterization of Musical Phrases," *Amer. J. of Psychol., 47,* 1935, 624-643.

mountain slopes, grazing cattle, and cozy farms across the river from Hanover.

Common observation, the authoritative writings of critics and specialists in esthetics, and some psychological researches combine to argue that there is no uniform intellectual experience in music; given selections do not cause auditors to imagine or recall the same "pictures" or to discern the same or even similar "messages" from the composers. The mental content of music is individual. Through performances of his score, a composer cannot expect to transmit any special thoughts to the listeners. Even program notes, sometimes in great detail, fail to force uniformity of thought content.

Attempts to discover the *representative* effects of music have led to negative findings along the line indicated above, but other investigations into the *expressive* effects have been more fruitful. While it is true, Gundlach reasoned, that our individual ideational responses depend upon our specific past experiences, current motivations and bodily conditions, it is still possible that a given musical selection can somewhat consistently determine the direction of our emotional moods. The *Pastoral* might evoke reasonably similar emotional trends in most listeners although the ideational expressions of the emotional trend could be expected to vary greatly.

Gundlach set out to investigate this hypothesis concerning the character of music. By character of a musical selection he means the emotional mood or temper, in general terms, which is aroused by a composition. Since musical composition is highly conventionalized, with many standard symbols and even clichés, he proposed also to discover the relationships between the experimentally determined character of musical selections and their compositional structure. From a methodological point of view Gundlach's approach is interesting because he combined the techniques of the cultural anthropologists (ethnologists), the experimental psychologists, and the musical analysts. Most problems in modern psychology require ability to work effectively in several fields, or at least demand that the

investigator work jointly with competent colleagues in related fields.

I. THE ANALYSIS OF INDIAN MUSIC

Gundlach began his search for the "signs" of musical structure which consistently produce given emotions with an analysis of Indian music from six tribes. He proposed to relate the *musical structure* of Indian songs to the *emotional moods* of the songs by analyzing songs which were already expertly classified by ethnologists. He proposed also to study songs of similar emotional purpose from several tribes to discover whether tribal tradition or some more fundamental factor governs the relations observed.

Why use Indian music? Apart from an interest in Indian music for itself, several considerations suggest that the first study in this larger problem of mood-character might well be done with Indian music. In the first place, Western music is too complex, involving harmonic as well as melodic structures, and is not so easily classifiable by moods since it is less related to everyday living. Again, folk music is not suitable because the same melodies have been used to convey a variety of moods, as, for instance, the numerous verses in long ballads. A melody cannot have much mood-characterization if it can convey the whole gamut of emotions. Finally, the Indian songs of North America have the advantage of being closely connected with important aspects of life and are, hence, sung on definite occasions which have specific emotional significance. The songs are brief and are not used for more than one kind of situation. The words are usually less important than the melody. Further, songs of similar mood-classification were available from several tribes.

The songs. Over 300 songs recorded by ethnologists of the Smithsonian Institution were made available to the worker. These were from five tribes: Chippewa, Dakota Sioux, Mandan and Hidatsas, Northern Utes, and Papagos. Songs from the Ojibway tribe were also available from another source. For

comparative purposes, Gundlach made a smaller study of 53 folk songs.

Analytical procedures. The 300 Indian songs were classified as to *situation* in which they were sung. Since Indians use a given song only in its appropriate situation, this method was tantamount to classifying them according to emotional mood. Adequate samplings existed in the material for three broad classes: war songs, healing songs, and love songs. Other kinds of songs, such as game songs, gambling songs, and dance songs, were omitted from the study. The sub-divisions of the war and love songs, indicated in column 1 of Table 39, are self-explanatory.

The analysis of the musical structure of these songs included the following items:

1. Pitch of lowest note.
2. Pitch range in terms of number of whole tones from the highest to the lowest notes.
3. Speed, the number of notes of a given time-value per second, stated in metronomic markings such as are found in standard music.
4. Interval size, in three classes: intervals larger than a "musical third," major and minor "musical thirds," and intervals smaller than "musical thirds."
5. Rhythm, defined in three categories: even (notes of same length and pattern repeated), uneven (each pattern in complex time but the patterns repeated), and rough (each pattern irregular and patterns not repeated).

Results. The essential results of this study are contained in Tables 39 and 40. Table 39 reports the average pitch range, the lowest pitch, and the tempo of the songs in each of the classifications and sub-classifications for all the tribes lumped together. Table 40 presents the proportional frequencies, in percentages, of the three sizes of intervals and the three classes of rhythms for the classified songs of all tribes. In the original report further analyses are given of the significance of the differences which are found between different classes of songs in these musical variables.

The most clear-cut facts from Table 39 are:

1. Organization and general war songs are fast (91.5, 95.8), low (4.00, 3.81), and wide in range (10.61, 10.27).

2. Healing songs are quite slow (80.0), moderately high (5.96), and narrow in range (8.28).

3. Love songs are high (6.25, 7.79), except for sad love songs (5.26), slow (75.0 to 80.0), and medium range (9.50 to 9.93).

4. Battle, victory, and scout war songs are similar to love songs except that they are lower in pitch.

TABLE 39

THE AVERAGE PITCH RANGE, PITCH OF LOWEST TONE, AND SPEED OF SONGS IN EACH CLASSIFICATION FOR ALL THE TRIBES

	Range		*Pitch of Lowest note* *		*Speed* †	
Classification	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
War Songs						
Organization	10.61	1.73	4.00	2.93	91.5	16.87
Warpath	8.85	1.68	5.80	3.65	85.5	16.67
General	10.27	1.61	3.81	3.10	95.8	20.20
Scout	8.83	1.86	4.60	2.22	79.0	22.45
In battle	9.92	1.94	4.54	1.98	73.5	15.62
Victory	9.00	1.88	4.52	2.80	85.9	19.70
Recitative (great deeds)	9.53	1.41	6.23	2.64	83.7	17.46
Healing Songs	8.28	2.07	5.96	3.03	80.0	19.85
Love Songs						
Sad	9.62	1.77	5.26	3.66	75.0	15.18
Happy	9.50	2.14	7.79	4.27	80.0	13.50
Others	9.93	2.04	6.25	3.20	77.8	19.82
Average	9.46		5.30		84.0	

* Substitute pitch A for 1, A-sharp for 2, B for 3, etc.

† Some Ojibway songs omitted from Speed measurements due to technicalities.

The most clear-cut facts from Table 40 are:

1. Size of interval does not seem to be a differential between songs. Most intervals are small anyway, melodies tend to ascend and descend by small steps rather than by vigorous "pitch-leaps."

TABLE 40

THE PROPORTIONAL FREQUENCY OF THE DIFFERENT INTERVALS AND THE TYPES OF RHYTHMS FOUND IN THE VARIOUS CLASSES OF SONGS AMONG ALL THE TRIBES

	Type of interval (per cent)			*Type of rhythm (per cent)*		
Classification	*Large*	*Thirds*	*Small*	*Even*	*Uneven*	*Rough*
War Songs						
Organization	7	22	71	30	30	40
Warpath	8	17	75	23	41	36
General	8	19	73	28	39	33
Scout	14	17	69	0	31	69
In battle	11	23	66	16	22	62
Victory	8	21	70	25	24	51
Recitative	6	19	75	31	23	46
Healing Songs	9	14	77	40	40	19
Love Songs						
Sad	15	20	65	42	43	14
Happy	9	20	71	39	32	29
Others	15	20	65	27	50	22
All Songs	9	19	71	31	34	35

2. Rhythmic structure does bear a relation to emotional mood. (a) War songs have more rough rhythms than healing songs and love songs. (b) The stalking scout's mood permits no even rhythms. (c) Songs depicting battle action likewise have few even rhythms. (d) Healing and sad love songs show the greatest proportional number of even rhythms; obviously they have much in common in mood. They differ considerably, however, in size of interval employed. (e) Happy love songs have twice as many rough rhythms as sad love songs. (f) In all songs, regardless of purpose, there is about an equal distribution of even, uneven, and rough rhythms.

In the original report another column was presented in Table 40 which demonstrated that tribe-to-tribe differences in these songs were slight. Because of the technical complexity of the analysis behind the column of figures the details are omitted here, but the reader can accept the fact that Gundlach accomplished an important aim when he was able to show that, apart

from a few genuine exceptions, the relation of musical mood (as defined by situations) to musical structure goes beyond local tribal traditions.

Some readers may feel disappointed that Gundlach failed to find more concrete and non-overlapping variables which would clearly differentiate one type of song from another. Some might feel that psychologists cannot get anywhere in the field of esthetics and other complicated social phenomena because they are unable to isolate the variables with sufficient accuracy to permit simple one-to-one relationships between the observed responses (in this case emotional) and the stimuli which produce them. Psychology always will be faced with this situation. No social situations are simple. There is no hope that the analysis of emotional moods in terms of musical structure will ever be as simply stated as Galileo's tower experiment. The variables are too many.

Gundlach was aware of this problem and presented the following argument. A total picture of the songs warns us that no particular item can be picked out and labeled as the typical characteristic of a certain class of songs. One can readily see that the characteristics analyzed here are not component parts of the songs, but aspects holding simultaneously for every phrase of a song; and that the effect of one or another characteristic depends not simply upon itself, but also upon the other characteristics. The rhythmic units composing sad love and healing songs are quite comparable in isolation. But the emphasis in the healing songs upon small intervals, narrow range, and greater speed makes the rhythmic structure seem quite different. Again, the rhythm analysis of the gay love and the first group of war songs (organization, warpath, general) and the analysis of the types of intervals would indicate that they were highly similar songs. But the greater speed, range, and depth of the war songs change even the apparent character of the analytically similar aspects. Of course the words, facial expressions, gestural intonations, and loudness of the song as performed are additional factors of importance which cannot be taken into this account at all.

In short, Gundlach argues, we must not believe that an analyzed musical variable is the same in isolation as it is in context. An understanding of the laws of perception which the Gestalt school of investigators has demonstrated will be helpful on this point.

A further observation is pertinent. Gundlach points out that the possibility of differentiating these emotionally classified songs into clear-cut structural groups is reduced by the grossness of the original classifications. Although tribe-to-tribe differences were found generally to be slight, they were not absent. For instance, the Papagos are a pacific tribe and fight only in retaliation. Any Papago who kills must be purified. Their war songs, situationally defined, certainly have a different emotional slant from those of other tribes who might take some delight in a good killing expedition. Also, some Indian songs are personal to the singer; only he or those to whom he gives the right may use them. Other songs are tribal. Some are ancient; others have appeared since Western culture has impinged upon that of the Indian. In order to get a sufficient sampling, Gundlach could not fractionate his classification further. He points out, for instance, that lumping of varieties of happy love songs may lead to crudities which will reduce the differentiation of the songs in terms of musical structure.

Folk songs. The small section on folk songs can be dismissed briefly by saying that the investigator found results with respect to war and love songs which were essentially similar to those he found in the Indian material. The sample was small and is not reported in sufficient detail to permit a critical evaluation of the data.

Summary. Working with a collection of Indian songs which could be classified with considerable certainty as to the emotional moods which were expected in the listeners, Gundlach demonstrated that there are some definite relationships existing between what he calls the mood-character of the music and the musical structure. The problem is complex; no simple one-to-one relationships are to be expected. Though only exploratory, the research is a distinct contribution to the psychology

of music in general and also to the special problem of emotional responsiveness to music.

2. MOOD-CHARACTER IN WESTERN MUSIC

In Gundlach's Indian study he stressed the advantage of working with music which is only melodic and not harmonic. However, apparently, he still wished to tackle the problem of mood-characterization in relation to musical structure within the music of our own culture. He laid out a research which paralleled the Indian study as far as possible. Three problems were set:

1. What is the consistency and uniformity with which observers can report the emotional quality characteristic of a composition? He could not start with labeled works such as he found ready-made in the Indian ethnological materials. He had now to deal with such neutral musical titles as Chopin's *Fifteenth Prelude.* Therefore, the first task was to find whether a given selection of music, or some phrases of a larger piece, will evoke in a group of typical listeners emotional moods which are identical or at least similar from listener to listener. As a matter of fact he tried to discover whether the listeners could agree on what mood the composer was trying to produce. If this part of the research should prove negative, he could not go on with the study of structure.

2. If the mood-characterizations of pieces of music are found to be quite stable and identifiable, what compositional factors, what musical mechanics, are associated with the defined moods? This part of the research parallels the Indian study and is possible only if positive findings are obtained on the first problem.

3. What interrelations exist between various mood-characterizations? In a list of terms expressing the sundry musical mood-characters are all of the terms unique or do some moods go together or overlap? This is a matter of a factor analysis. The data are rather inconclusive in this part of the research and no further summary is to be made here.

Procedure. Forty fairly diverse pieces of electrically recorded music were selected, including piano, violin, orchestral, and other music from Handel, Bach, Debussy, Chopin, Gershwin, Herbert, and others. Obviously the experiment would have been excessively long had the listeners been required to hear forty full concert recordings. Instead, to make an hour-long experiment, the researcher used the first phrase of each number.[2]

There were 112 subjects who were tested in small groups of 20 to 40 persons. Of these 112, 33 were untrained in music, 25 had some training in instrumentation but not in theory, and 54 were advanced students in music courses.

The subjects were given recording sheets with 17 adjectives (which are given alphabetically in Table 41) listed after each number to be played, together with 8 blank spaces for writing in other words. The subjects were told to listen to the phrase so as to determine what mood or attitude the composer had succeeded in expressing. Note how the emphasis was on getting away from personal memories for an objective judgment of what mood was being produced. They were to check appro-

[2] Reliability might have been increased had longer passages been employed, but then, on the other hand, if the whole selection had been played numerous mood-characterizations might appear, even quite contradictory ones. For instance, the typical storm and strife followed by quietness type of musical theme could hardly be characterized as being either stormy or quiescent. All we can say is that Gundlach's listeners listened to the initial phrases of the performances. All through this research the reader must be mindful that the moods which these given pieces are said to induce apply only to the introductory phrase of each. Of course, this yields significant data since the musical analysis which followed also was limited to the initial phrases. Since he was using only phrases, it is worth asking why original compositions should not have been used so as to rule out the recall of past experiences which are associated with recognizing the opening phrase as belonging to an "old musical acquaintance." We wonder what feelings of incompleteness there were when particularly loved opening passages were suddenly stopped. Answers to such questions must await further research, but are asked to suggest to the student the complications which appear in investigating personal reactions to musical passages.

priate adjectives (more than one, if necessary) and to add other terms if they felt the need of them.

There is some question as to how these 17 adjectives were finally selected. Apparently, Gundlach had expert help to prepare a list of terms which would be acceptable to musicians and likewise would make sense to the experimental psychologist. He does report that several "dimensions" of terms were used. One set pertains to the dynamical aspect of music, terms like brilliant, animated, uneasy, tranquil, dignified. A second set refers to emotionally toned attitudes, such as triumphant, exalted, glad, somber, melancholy, and mournful. The third dimension is tentatively identified as referring to playful self-consciousness, including these terms: delicate, whimsical, flippant, sentimental, awkward, grotesque.

Results: *Consistency in relating categories of moods to the musical phrases.* Anyone who has studied in the field of emotions recognizes the "bug-bear" of using names to describe emotional experiences. The old exercise of matching emotion-names with facial expressions comes to mind. Gundlach did find considerable uniformity of response from his listeners. Twenty-four times at least 50 of the 112 listeners agreed in ascribing a given mood-name to the initial phrase of some particular musical selection. Other adjectives were assigned less frequently to the music. For some music no term was consistently used; no mood-character could be indicated for such selections.

For instance, the Bach *Brandenburg Concerto, No. 2,* received this vote: 66, glad; 51, brilliant; 50, triumphant; 37, animated; 10, exalted; and only four of the 17 terms were not applied at all. There were some clear-cut agreements on this selection; it is glad, brilliant, and triumphant in character. Mozart's *Symphony in G* received 29 votes for glad, 26 for uneasy, and 16 for whimsical, while ten other mood-categories were mentioned.

By using the method of rank-difference correlation, the investigator was able to conclude that there was great similarity between the responses of the untrained and the trained sub-

jects in the selection of appropriate terms for the description of the various pieces. Of 17 coefficients of correlation,[3] 14 ranged between .70 and .91, and three were lower. Further, using the split-half technique of randomly dividing his sample of listeners into two groups, he found the reliability coefficients were all over .75, except for the use of the term *exalted.* We can conclude, then, that while musical experience does make some difference, the judgments of the character of the music depend more on the nature of the piece and less on the training of the listeners.

It was observed that some mood-terms were used more frequently. The rank-order of the 17 terms was: animated, uneasy, tranquil, dignified, glad, somber, brilliant, whimsical, melancholy, grotesque, mournful, delicate, sentimental, triumphant, exalted, awkward, and flippant. Animated was checked 581 times, flippant, 167 times. Gundlach observes that high in the list are those terms which he originally selected as being dynamic expressions.

While the reader may have hoped that Gundlach would have demonstrated more unequivocal association of certain mood-terms with certain pieces of music, he should be reminded that the method here employed was more fruitful than those used in earlier studies, that the technique probably can be improved (for example, by longer musical exposures), that our language for describing emotions is extremely inadequate, and that emotional experiences themselves overlap each other considerably.

Results: *Relation of mood-character to musical mechanics.* As in the Indian study, Gundlach made an analysis of the

[3] In determining these correlations Gundlach first counted the number of times that each descriptive phrase, e.g. "brilliant" had been used for a given musical selection by the expert raters. The most frequently used term was then ranked first, the next highest frequency second, etc. When the rankings by the expert groups were compared with those of the untrained group by the Spearman rank order method of correlation, it was shown that in most cases the two groups of judges agreed quite closely in the selection of the most fitting terms for the description of each selection.

musical structure of the selections in order to relate such mechanics of composition to the categories of mood-character which had now been ascribed to the selections.

The analysis was along seven lines. (1) Loudness, as judged by musical experts. (2) Tempo, as judged by experts and expressed metronomically. (3) Pitch range of the melodic line, that is, the number of whole-step intervals from the highest to the lowest notes. (4) Pitch range of the orchestral accompaniment. (5) The pitch of the mid-most tone in the melodic line, which would give a rough measure of the highness (or lowness) of the selection. (6) Sizes of intervals, tallied much as the Indian intervals were counted. (7) Rhythms, using the even, uneven, and rough categories as previously.

The technical details of this analysis are too cumbersome for complete report here. Gundlach points out that not all the comparisons are clear-cut. Statistically many comparisons are unconvincing. But, taking a conservative view, he has constructed summaries of the elaborate tables which give the gist of the findings. Table 41 is constructed from several of his exhibits.

As with the Indian music, he found that tempo and rhythm are the most significant discriminators of mood. Likewise, stress must again be placed upon the fact that no single element of musical mechanics accounts for a good characterization, but rather, a musical mood is the resultant of a combination of variables, a Gestalt or configuration. Thus, music characterized as grotesque or uneasy is fast and rough in rhythm, while music which is said to be brilliant or glad is fast but with many smoother rhythmic patterns. Similarly, somber and sentimental moods are both found to be associated with slowness and unevenness of rhythm, but differ in that somber music is low in overall melody, while sentimental music is higher.

Summary. Gundlach's research is essentially a preliminary excursion into the intricate problem of discovering whether there are certain musical structures (tempo, rhythm, and the like) which act as mood-determiners in musical listeners. In

Table 41

MUSICAL STRUCTURAL CHARACTERISTICS FOUND TO BE MOST CONSISTENTLY ASSOCIATED WITH THE JUDGED MOOD-CHARACTER OF THE MUSICAL SELECTIONS *

Mood-category	*Tempo*	*Rhythm*	*Interval*	*Pitch level*	*Pitch range*
Animated	fast	few uneven		high	wide
Awkward			many 1sts, 2nds		narrow
Brilliant	fast	many smooth few uneven			wide
Delicate	slow	many uneven	many large		narrow
Dignified	slow	many uneven		low	narrow
Exalted		many uneven	many large		
Flippant	fast	few uneven many smooth			
Glad	fast	many smooth	many large	high	wide
Grotesque	fast	many rough few uneven		low	wide
Mournful	slow		many 1sts, 2nds	low	narrow
Melancholy	slow				
Sentimental	slow	many uneven		high	
Somber	slow	many uneven		low	narrow
Tranquil	slow			low	narrow
Triumphant			many 3rds		
Uneasy	fast	many rough	many 1sts, 2nds		wide
Whimsical	fast			high	

* Gundlach did not summarize the data in terms of loudness and apparently lumped melodic and orchestral range into one group in the tables from which Table 41 is made.

fact, it is even incomplete with respect to the question of whether there are any stable mood-responses which people can make in the presence of certain musical productions.

The data do give us a preliminary answer which argues for more research and suggests that his method, with improvements and extensions, may become very fruitful. It surely argues that there is much to be done in the field of psychology of music under the joint sponsorship of experimental psychologists, scientifically minded musicians, and anthropologists.

Finally, we must ask ourselves about the causes of the relationships which have been observed. By this time the reader is sophisticated enough in psychological understanding to be unsatisfied with a suggestion that people "just naturally" have certain emotional responses when musical productions are heard. It is quite likely that any explanation in terms of biological inheritance of specific modes of response will be unacceptable. The answer must be sought in the culturalization process of the individual. Two questions then arise. How does the individual person learn to respond emotionally to his particular musical culture? How does the society as a whole come to associate certain emotional responses with particular kinds of musical structures?

The reader will find the answer to the first question when he reads about the psychology of learning, about the growth of language, and about the socialization of motives. Music is but one of the cultural hand-me-downs which children learn to understand in greater or lesser degrees, just as they learn to understand the language of their society in greater or lesser degrees. The same broad principles of growth and learning apply.

The second question causes us to seek an answer in the various theories of the origin of music. These cannot be discussed at this point, but one particular item should be raised. It is that much of our musical experience, if not most of it, is associated with muscular activity. Sitting quietly at a concert is only an isolated aspect of music and even there we do not really sit still. Primitive peoples experienced most of their

melodic, rhythmic, and harmonic pleasure while working, while playing, while rocking babies to sleep, in all sorts of motor activities. While engaged in these activities emotional responses, apart from musical behavior, were aroused. So it is probable that the low-pitched, slow, narrow-ranged, uneven rhythm of the saddened mother while watching over her dying baby is a direct consequence of her total physiological condition at that time. Since such motor patterns and accompanying emotional experiences came to have meaning, man began to employ these particular sounds and motor gestures when he wished to convey musically the emotion which he had in mind. This short-circuited statement of the dependence of musical emotion upon the total experience of man, especially upon his muscular activity, is meant just to stimulate the interested student and obviously does not answer all of his questions.[4]

[4] The student will find C. M. Diserens' volume, *The Influence of Music on Behavior,* Princeton University Press, 1926, of considerable interest.

CHAPTER 33

Physical Education and Recreation

FIFTY years ago few schools and colleges had departments of health and physical education. Few cities sponsored playgrounds and other recreational centers. The profession of physical education did not exist. Broken-down wrestlers at the one end and medical men with a flair for student health and athletics at the other extreme filled the few positions available. Full-time coaches were not recognized as members of the faculty and many coaches in colleges did their coaching as a sideline.

Today nearly every school and college has a department of health and physical education. No longer are these functions generally considered as extra luxuries; they are accepted as part of the curriculum. Most cities have at least caught the gleam of the "playground movement" and even rural recreation is becoming organized into a program. The federal government is vitally interested in recreation especially with reference to the effective use of national parks. From the cradle (nursery schools) to the grave (octogenarians in Florida and California) people are deriving the physical and mental hygienic benefits which arise out of sports participation, both team and individual, and other forms of recreation.

Psychologists have contributed to this emphasis on physical and mental health through stressing effective use of leisure time. Here we shall limit ourselves to the experimental phases of their contributions, but it should be remembered that psychologists also have shared in the development of the philosophy and actual practice of modern recreation, mainly through an interpretation of known laws of psychology. Much that has been said and written about the character-building value of

athletic participation has, in all probability, been overstated and needs experimental verification.

In this section we shall review just two studies which sample the psychological research in this field. First, there is Miles's application of the well-known techniques of measuring reaction time to the practical situation confronting the football coach who wants to select linemen who are fast in getting away when the signal is called. Dr. Miles brought standard laboratory techniques right out on the playing field.

Second, Lehman and Witty employed the questionnaire technique to study the play interests of children, a problem which has considerable bearing on any plans to set up psychologically correct programs of recreation. This report is included partly to bring before the reader a carefully set up questionnaire study which went beyond the mere "nose-counting" with which so many census-type questionnaires end.

There are numerous other investigations which merit inclusion in this section, some of which have been done by psychologists and others by men in other fields, such as physical education itself. Much of the work comes under the rubric of educational psychology and much under the classification of tests and measurements. The role of recreation and play in the therapeutic programs of clinics and hospitals is just beginning to command attention. Much "practical psychology" on personal efficiency is available. The field is wide open for alert research-minded persons who wish to concentrate their efforts on athletics, physical education, and recreation.

Finally, attention should be called to the fact that not all recreation is physical or athletic. The whole psychological literature on music, art, drama, hobbies, and manual arts bears on the modern "curriculum" in recreation which a community might try to institute. Further, since most of the recreational emphasis is on the childhood and adolescent periods, the psychological contributions of specialists in these age groups need attention. Recently there have been some attempts to set up controlled camping situations, out of which experimental materials of considerable merit have emerged. In brief, opportuni-

ties for research, both "pure" and "applied" in the fields of athletics, physical education, and recreation, are unlimited and uncrowded; possessing a distinct lure for the next few decades of social development in this country, these fields should command some of the keenest psychological workers.

1. REACTION TIMES OF FOOTBALL PLAYERS [1]

Moving the reaction time experiment out of the laboratory to "real" situations. In the typical setting for laboratory study of reaction time, the subject lifts his finger from a circuit-breaking key whenever he perceives a prearranged signal. The stimulus and reaction situation can be very simple or very complex depending upon the problem being studied. Description of typical experimental procedures and their meaning are given in standard experimental manuals and texts.[2] For some time now psychologists have been interested in bringing these laboratory techniques into more fruitful use in connection with problems of industry, highway safety, and athletic ability.

The swimmer poised on the edge of the pool, the track mar. on his mark, the football lineman awaiting the "hike" are all about to demonstrate significant individual differences in reaction time. Their muscles are in a state of readiness to react quickly when the signals are given. The sprinter who "jumps the gun" when the signal is delayed or when some extraneous noise in the stand or gallery occurs and "offsides" in football illustrate the condition of high neuromuscular tension which is found in these realistic reaction time situations. Even if the lineman is unaware of the precise physical laws which are

[1] Adapted from W. R. Miles, "Studies of Physical Exertion," published in three parts. Part I, "A Multiple Chronograph for Measuring Groups of Men," *Amer. Physical Educ. Rev.*, XXXIII, 6, 1928, 379-387. Part II, "Individual and Group Reaction Time in Football Charging," *Research Quarterly*, II, 3, 1931, 5-13. Part III (with B. C. Graves), "Effect of Signal Variation on Football Charging," *Research Quarterly*, II, 3, 1931, 14-31.

[2] For a historical and a general experimental review of reaction research, consult Ch. 9 in H. E. Garrett, *Great Experiments in Psychology*, Century, 1930, and Ch. XIV in R. S. Woodworth, *Experimental Psychology*, Holt, 1938.

operating, he does know that a quicker getaway helps bring him to his peak charging efficiency quicker, and if quickly enough, to the detriment of his opponent.

For the reader's general information, let us point out some important criteria of a good lineman's charge. (1) He must charge before the opposition moves; when he is on the offensive he has the advantage of the signals. (2) He should be a fast charger even when he is on the defensive; that is, he should react quickly to the cues (snapped ball and opponent's charge) which are available. (3) He must mobilize himself well with respect to strength and direction of charge. (4) He must move with the team as a whole in whatever pattern of movement he has been assigned. Miles did not investigate all of these factors. His main work was on item 1; there was also some analysis of reaction time in relation to offsides, which is a part of factor 4.

Measuring the reaction times of several persons at once. Miles noted an important variable which had to be brought under control, namely, that in group games the interplay of competition and mutual interstimulation are very essential parts of the performance. Miles's first task, then, was to construct a reaction time apparatus with which he could measure several men at the same time, such as the whole line in football.

The portable multiple chronograph [3] is pictured in Fig. 56. Its essential features were:

1. A motor-driven drum, 1000 mm. in circumference, covered with a "skin" of paper over a fine mesh, provided the recording unit. The rate of rotation being one revolution per second and constant, Miles could measure reaction time if the moment of the starting signal and the moment of the reaction of each man were impressed in some way on this paper.

2. A falling-ball technique of marking the starting and reaction times on the drum was employed. Eight latches were mounted above the drum. Each latch mechanically held a golf ball under very slight pressure which could be released by the

[3] Time-recording instrument. This particular type is now superseded by much simpler electric stop-clocks.

pulling of a lever mechanically or by magnets. Usually strings were attached from the releasing latches to the special vertical triggers which the athletes pushed. The balls were equal in

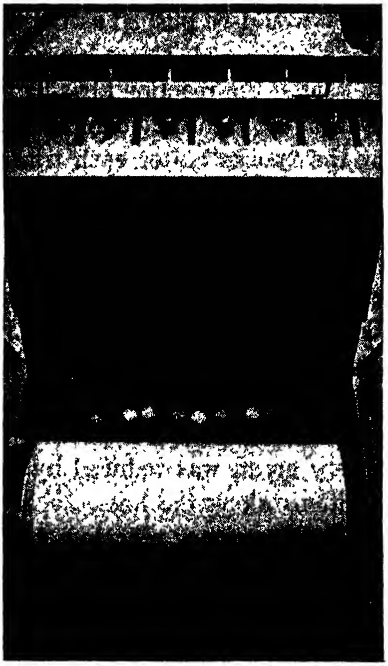


FIG. 56. Instrument is the Miles Multiple Chronoscope for measuring changing speed of seven football linemen.

Balls are released by string tension on levers near top of picture, dropping to make marks shown on paper over wire mesh-covered drum run by motor. The extra ball is released by the experimenter at the same time as the charging signal is given. (Courtesy of W. R. Miles.)

weight and fell equal distances. Upon being released the balls fell on the paper, "pocking" its surface just once since the latches were arranged so that the balls would bounce off at an angle into a net. The reaction times were the distances in mm. between the "pock" of the signaler's ball and the "pocks" of the balls released by the subjects, the distance being converted

into milliseconds (thousands of a second, commonly abbreviated *ms.*).

3. The reaction keys, consisting of hinged bars connected by strings to the latches, were placed before the heads of the linemen so that they would be pushed as the men charged forward at a predetermined signal. These hinged boards could be rigged as electric keys if the reaction keys had to be placed at greater distances from the recording equipment.

4. By marking the "pocks" after each trial with trial numbers and initials of players, several trials could be recorded on one paper on the drum. Replacement of the paper was simple. A special ruling device aided in accurate measurements.

5. Limitations of the equipment introduced experimental errors said by Miles to be not over 1 per cent. First the latches could not be perfectly matched for tension. Second, some of the latch-releasing strings had to go out at wider angles from the latch rack since a football line is about 18 feet long. This error, while not very great, was reduced by rotating the men in the several positions along the line.

Procedures for measuring reaction times of men in group activity. The multiple chronograph was placed at the side of the practice field. An assistant coach worked with the men in groups of seven, usually groups which had been practicing together. They stood in line in their regular stances with their heads against the vertical triggers. The football stance on the field is designed more for force of action than for speed. It involves large muscle action rather than refined movement of one or two small muscles. The Miles apparatus required a four-inch forward motion for the releasing of the latches in order to avoid accidental release by the small motions that are inherent in the very energetic acts of getting ready for a charge on the signal. The experimenter served as a "quarterback." In the first study, the men were to charge forward ten to twenty feet on the "hike" of the sequence, "position—signals—hike"; one to three seconds elapsing between signal words. After each trial, the men gathered around the apparatus to observe their approximate reaction times as Miles was marking the trials

and initials on the "pocks." The motivating effect of this procedure is apparent. To avoid "place errors" in the line-up, the men rotated positions. The seven-men squads were given seven trials. In the second unit of research variations of signals were used, the subjects being informed of the signal in what would correspond to a "huddle."

Subjects. For the study of individual differences in reaction time of football players on the line, 87 Stanford University players were measured by the apparatus. This included 53 out of 55 men who were listed as the official squad. In the study of the effect of signal variations only four men were employed for an intensive study. They were considered typical.

Auditory reaction time in the laboratory and on the field. In the laboratory it is known that, typically, practiced adults can react in from 120 to 170 ms. when a sharp auditory signal is given. Unpracticed adults, according to Miles, will need about 180 to 250 ms. Before studying the reactions of his subjects in regular football stance he had a few of them react by facing away from the apparatus and pulling the strings which were held over their shoulders. Their reaction times in this gross movement of pulling a string were from 203 to 272 ms., averaging 238 ms.

For 87 players, Miles found an average reaction time of 389 ms. Individuals ranged from 268 to 523 ms. The average of the 87 subjects' average deviations (A.D.) was 39 ms. from their individual mean reaction time.

In summary, the football players, on the average, required about .4 second to get off their line on a charge, but the slowest required twice as much time as the quickest—the range of times being 268 to 523 ms.

Reaction times of special groups. The eleven men in the starting lineup of the "big game" of the season had an average reaction time of .353 ms. which is 28 ms. shorter than the average of the 53 men who constituted the entire official squad at that time. Two very slow men, 431 and 434 ms., were among the eleven, apparently because of other qualities than speed.

The 23 men who were dropped from the squad or were not listed later on the official squad had an average reaction time of .392 ms. It should be made clear that the coaches knew nothing of these men's reaction times; their scores were not the basis for exclusion or inclusion on the team.

In Table 42, the most important and suggestive sub-study is summarized. The individual groups are relatively small but the positions rank themselves in an order that is not illogical.

TABLE 42

CHARGING SPEED OF VARSITY FOOTBALL PLAYERS ACCORDING TO POSITION

N	*Position*	*Average*	*Ave. dev.*
18	Backs	360 ms.	38 ms.
9	Ends	377	34
10	Guards	383	42
11	Tackles	395	41
5	Centers	444	30

Space does not permit a summary of an attempt to relate coaches' ratings of men for speed and efficiency to the objective measures of speed. The computed relation was not high, due to three factors. First, the known fact that ratings made as crudely as these are not highly reliable. Second, Miles's data were based on too few cases and too few trials per case to permit refined individual comparisons. Third, as ought to be apparent to the reader, speed of reaction is only one of the many factors determining the success of a football player.

The problem of signals. The successful team not only demonstrates speed but also shows unison of movement. Snappy, precise, rhythmic movement of the whole team is the objective towards which the training is directed. Signals are called to coordinate team action in every play. This raises the question as to what type of signal-calling is most effective. Naturally, the signal must be one which misleads the opposing men; it must, therefore, contain an element of surprise. The offensive line and backfield must not reveal through any cues their exact split-second of charging. To settle some of the controversial

issues on signal-calling raised by experienced coaches, Miles and Graves studied four players intensively. The questions were:

1. What rate of calling digits in the signal is optimum? Seven-digit signals were called at rates of 40, 60, 100, and 120 digits per minute, as governed by a metronome.

2. Are rhythmic signals or non-rhythmic signals better?

3. Are anticipatory or non-anticipatory signals better? The cue for charging could be any number called after the first two; seven digits was the limit. In the *anticipatory* signaling, the men were told on a given trial, for example, to react to the fourth digit called, regardless of what the digit would be. In the *non-anticipatory* signaling, they were told to respond only to the digit "two" wherever it appeared between the third and the seventh digit in the signal.

The data were analyzed in terms of reaction time; offsides, as recorded by latches being pulled before the proper cue; and unity of charging of the four men, as measured by deviations of each man from the average of the four men.

The signaler spoke into a voice key which would release the starting-time latch on the call of the proper signal. A voice key is an electrical switch operated by the sound of the voice.

For lack of space this review is confined to the effects on reaction time. Details of the study of offsides and of unity are omitted but summaries of the conclusions are given. Tables 43, 44, and 45 give the essential data.

TABLE 43

EFFECT ON THE SPEED OF CHARGE OF CALLING "ANTICIPATORY" SIGNALS AT DIFFERENT RATES

Rate per min.	*Number position in series* 3d	4th	5th	6th	7th	*Ave.*
40	210 ms.	119 ms.	170 ms.	153 ms.	207 ms.	171 ms.
60	173	102	152	104	183	143
100	107	73	99	85	78	88
120	104	118	112	90	106	106
Ave.	148	103	132	108	143	127

TABLE 44

EFFECT ON SPEED OF CHARGE OF CALLING "NON-ANTICIPATORY" SIGNALS AT DIFFERENT RATES

Rate per min.	Number position in series 3d	4th	5th	6th	7th	Ave.
40	473 ms.	470 ms.	415 ms.	431 ms.	467 ms.	451 ms.
60	460	418	424	420	417	428
100	411	397	392	400	383	397
120	443	378	401	485	433	428
Ave.	447	416	408	434	425	426

TABLE 45

SPEED OF CHARGE IN RESPONSE TO SIGNALS GIVEN ARHYTHMICALLY

Kind of signal	Number position in series 1st	2d	3d	4th	5th	Ave.
Antic.	470 ms.	459 ms.	449 ms.	443 ms.	441 ms.	452 ms.
Non-Antic.	640	538	460	498	478	523

Rate of calling signals. In Table 43 it is seen that a rate of 100 digits per minute is optimum for calling anticipatory signals. In Table 44, it is seen that a rate of 100 digits per minute is also optimum for calling non-anticipatory signals. In tables not shown here, it appears that for anticipatory signals a rate of 100 digits per minute also results in optimum unison of charging, that is, in less man-to-man variation of charging; but for non-anticipatory signals the slower rate of 60 digits per minute is better with respect to unity of response.

Anticipatory versus non-anticipatory signals. The values in the right-hand columns of Tables 43 and 44 show that the reaction time to non-anticipatory signals (a certain *number* to be waited for) is about three and one-half times as long as to anticipatory signals (a certain *digit-place* to be reacted to), 426 ms. as compared to 127 ms. Analyzed, it is apparent that in non-anticipatory signaling the player cannot react until he has perceived the number which has been agreed upon and he must avoid responding to the rate of calling and the rhythm; it is a very complicated mental act. In the case of anticipatory signaling, the tempo of calling actually becomes the main func-

tioning part of the signal and after two digits have been called the player can muscularly respond to the rate of calling and actually be in action more nearly simultaneously with the charging cue. Data show that this anticipation leads to more offsides penalties; there were seven times as many cases of offsides in the anticipatory trials in comparison with the trials in which a man did not react until he had cognized the signal *content* rather than the time of arrival of the signal.

Non-rhythmic signaling. In Table 45 data are shown which were obtained under the condition of both anticipatory and non-anticipatory signals but with the rhythm of calling the digits altered so as to be highly irregular. The values in the right-hand column are to be compared with those in the right-hand columns of Tables 43 and 44. Reaction time is slowed markedly; apparently the men could not "pre-organize" themselves. Note that the non-rhythmic calling of anticipatory signals cancels most of the advantage of such signaling over non-anticipatory ones and the average time is actually as bad as for rhythmic non-anticipatory signals.

Discussion. The importance of Miles's research consists mainly in his applying laboratory techniques to a life-like situation which is worthy of study both because of our interest in the basic problem of reaction and also because of our desire to apply known techniques to technological problems; in this case, to athletic coaching. If the study were being done today, new types of convenient electrical stop-clocks would replace his bulky, yet efficient, chronograph.

The study involves a recognition of the important variable of inter-stimulation. The isolated human being is always an abstraction, especially so in the highly emotionalized atmosphere of competition in modern collegiate football! Unfortunately we probably shall not be able to study the reactions of the men under the stress of real competition. However, it is suggested that a good extension of this study would be to repeat it with equipment for each of two opposing lines. Then we could study not only the men's responses in somewhat more game-like procedures, but also we would have data concern-

ing the reaction times of the opposing linemen who would have no information regarding the arrival of the signal other than the beginning of the quarterback's voice. This study would permit a study of the reaction time advantages the offensive team has over the defensive team. Doubtless some opposing men, even though uninformed on signals, might get away quicker than some of the men who knew the signals.

One major limitation of the first part of Miles's study was that he obtained only seven reaction-time records for each man. In the laboratory, fifty or a hundred trials are usually taken under each condition. The reliability of his obtained individual differences would have been increased with more trials. Anyone pursuing the problem of differences between players in general and between players in different positions on the team must keep this fact in mind. In the signaling study, an intensive procedure was used, the four men making 1440 trials. Increasing the number of trials per man would probably make more meaningful the attempts to correlate reaction time and coaches' ratings of their players.

Although it was beyond the scope of Miles's study, it would be interesting to investigate what other physiological and psychological variables play a role in football charging. We have no real experimental evidence, fully analyzed and verified, which demonstrates that speed of reaction is the most important factor. Likewise, it might be found that different patterns of abilities yield equal efficiency. From everyday knowledge of physics we know that a small fast-hurtling player can match the physical force of a heavier slow-moving man. How about the role of speed in different positions? The present data give several cues for extensive research. Someone should study the relation of stance to charging speed and efficiency. Also, it would be well to construct a setup which would measure both the reaction time of the player and his accumulated force, say, three feet ahead of his starting position.

Needless to say, the studies on signaling are vitally significant and ought to receive practical application. The offensive line which gets off first and as simultaneously as possible

clearly has the advantage over its opponents. By agreeing in the huddle to snap the ball and charge on the *fifth* digit, on a certain play, the team can get away in one-third to one-fourth the time needed to get away if they were agreed to wait for an arbitrary number, say, "six" to be heard from the quarterback; this fact is of great importance and ought to be capitalized. Of course, the larger number of offsides would need to be considered. Agreeing to respond to the fourth digit which will be a "nine," for instance, might even reduce the time still more, although the mind-set of "fourth digit" would be dominant.

The reader is reminded that the same general approach which Miles employed can be adapted to any sport wherein speed of get-away is important, or for that matter, to any sport where speed itself is important.[4]

A further question might be raised: "Granted that reaction time is important, why not measure it in the laboratory rather than on the playing field? Is not a person's reaction time the same for all situations?" The answer is "No." Experiments [5] show that reaction time varies with the total setting, the muscle groups involved, and the condition of the person. Reaction times under various conditions yield low coefficients of correlation. We have no reason to believe now that the fastest football lineman will also be the fastest man braking his car, warding off a falling brick, or shooting a gun at moving targets. In short, psychologists who wish to study the reaction abilities of athletes must devise apparatus and techniques for studying them in athletic situations, simulating not only posture and stance but also motivation, fatigue, emotion, and other important variables.

[4] W. R. G. Bender has summarized the work up until 1934 on the starting times of sprinters, in *Research Quarterly,* V, 1934, 1, 72-78.

[5] See reference in Woodworth, *op. cit.,* and, also, R. H. Seashore, "Individual Differences in Motor Skills," *J. General Psychol.,* III, 1930, 38-66.

2. CHANGES IN PLAY INTERESTS WITH AGE [6]

In this era of "the new leisure" we have more time for play. Leisure is no longer a luxury but a part of normal living for everyone. Society is organizing itself better for the supervision of these leisure activities. But play itself is not a new phenomenon; people have always played; our modern games have their roots in antiquity. Sharp distinctions, furthermore, cannot be made between our various daily activities. We may rest while we are recreating with a game of chess; we may work while planting an unnecessary garden; we may be enthusiastic about our vocation or profession to the point of calling it play. The same activity is play to one person and work to another; to the professional ball player, spring training is work.

While in earlier days, moralists attempted to link play with the devil, today we believe play is necessary for normal growth of the child and for continuance of effective living in maturity. Play is directly educative in effect; it is a mechanism of individual adjustment; it is at the root of mental hygiene. Play is a great socializing force. It is a worthy field for psychological investigation.

Theories of play. There are several theories of play. One is called the "surplus energy theory," the most elaborate statements of which were given by the poet Schiller and the philosopher Spencer. In brief, play is said to arise out of the need of the organism to discharge excess physical energy, while art arises out of the need to discharge mental energy. Couched in the "faculty" psychology, some of Spencer's writings do not seem to make sense to us now. There is no doubt, however, that play does consist in the expenditure of energy which is not needed for more necessary activities of living. This thumbnail review must remain incomplete and cannot do full justice to the complexities of Spencer's writings and those which he inspired.

[6] Condensed from H. C. Lehman and P. A. Witty, *The Psychology of Play Activities,* Barnes, 1927, 242 + xviii. Material is chiefly from Ch. V, "General Age Growth."

Karl Groos, around 1890-1900, proposed a "practice theory" of play. Children play in ways which give them practice in the activities which later in life are necessary for them. Girls play with dolls, boys drive imaginary trucks, and so on. The main criticism, as the reader quickly sees, is that much of children's play is not direct preparation for adult activities. Playing baseball is not rehearsal for banking or farming. Further, why do adults play? Groos added a "safety valve" (catharsis) notion to his larger idea; play can release pent-up emotion. With this latter aspect, mental hygienists will agree.

A famous theory of the early twentieth century was that of G. Stanley Hall, called the "recapitulation theory." He believed that during childhood the individual goes through the stages of evolutionary development and man's cultural history. "Ontogeny repeats phylogeny." Swimming repeats part of infrahuman behavior. In the "big injun" stage the pre-adolescent is reliving cultural history. Critics simply point out that Hall was right in asserting that individual development is determined by past organic and cultural history, but that the step-by-step recapitulation of evolution is not borne out by facts. Why, for example, should a boy like to swim and live in caves in the same six-month period and why do girls play with dolls in their first few years of childhood?

These and other theories of play are not so much invalid as they are limited and partial explanations of what play is like and for. John Dewey has argued, in a way which is more acceptable today, that a reasonable explanation of play can be sought in the fact that organisms are alive, active, and metabolizing. The problem for scientific study becomes, then, a matter of accurately describing the nature of play activities and observing the role of such activities in the lives of the persons in the whole process of socialization from birth on.

The problem of play interests. Defining play as activity in which the individual engages "just because he wants to" (omitting such physiological activities as eating and sleeping), Lehman and Witty set as their task a rather comprehensive investigation of the conditions under which various types of play

occur. Their goal, of course, was a practical one of helping psychologists, educators, and parents to understand better the play life of children, especially, and to suggest means for its effective control.[7]

We shall review their study of play with reference to the *age* variable only. Other analyses were of early childhood, sex differences, race differences, rural and urban differences, individual differences, season of year, school progress and play, mental ability and play, and the relation of play to educational and vocational guidance.

Method—the Play Quiz. The experimenters constructed a check-list (a special type of questionnaire) which was to be filled in by large groups of children. This check-list, called the *Play Quiz* when presented to the children, was a comprehensive list of 200 play activities. Some of these items are noted in the tables on later pages. The child was to check each activity in which he had engaged during the past week. Each child also indicated the three activities which had given him the most fun or which he liked best. Also, he marked the one on which he had spent most time. In parts of the studies, children also marked those in which they had participated alone.

The Play Quiz technique was designed to disclose: (1), the play activities most commonly engaged in by representative persons from five to twenty-two years of age residing in representative communities; (2), the play activities best liked by these individuals; (3), the games and other play activities consuming the greatest amount of time; (4), the extent to which a given child participates with other children in his play activities; (5), the effect upon play behavior of such variables as age, race, sex, season, intelligence, community, etc.

A careful investigator does not sit down at his desk and dash off a questionnaire. Good check-lists and free answer questionnaires are extremely difficult to make. Several methodological details will reveal the care used by Lehman and Witty. (1) After preliminary studies the number of activities was in-

[7] In Puritan days, to "control" play meant to forbid it. Today the term means to regulate by enhancing as well as by forbidding.

creased from 144 to 200 so as to include the most common activities which were "written in" by early subjects. (2) More common games were placed first to enable the child to get the idea of the list more quickly; the directions would seem more comprehensible if the first examples were almost invariably checkable by the children. (3) For smaller children the Play Quiz was filled out in two sittings. (4) Preliminary studies of various types of instructions to the teachers and children were made. (5) In two later studies, an additional item was included to reveal the extent to which children played alone or in groups. (6). Cognizance was taken of the need for studying the characteristic activities of the different seasons; the Play Quiz was given in November, February, and April. There probably should have been a summer sampling also.

Continuity and permanence of interests. Following Shakespeare's lead in his famous "seven ages," writers on childhood have for decades attempted to describe distinct ages of man. Typical of these classifications is that of Joseph Lee, founder of the American "playground movement."

1. First three years Creative impulse begins to manifest itself
2. Ages 3 to 6 ....Age of impersonation
3. Ages 6 to 11 ."Big Injun" age or age of self-assertion
4. Ages 11 to 14 ..Age of loyalty
5. Ages 14-21 . .. Apprentice age

This classification is only one of the many diverse ones, whose diversity proves their inadequacy. While such age-level divisions may seem useful in setting up practical programs, they are nevertheless dangerous if they cause us to warp our play programs for children into categories which have no reality.

Continuity of growth is the salient feature of childhood. Breaks between "levels" or "ages" do not exist. Even the onset of pubescence [8] is not accompanied by such sharp changes in activities, attitudes, and emotions as was once believed. Tables 46 and 47, constructed from Play Quiz data, show the essen-

[8] Lehman, H. C., and Witty, P. A., "A Study of Play in Relation to Pubescence," *J. Soc. Psychol.*, I, 1930, 510-523.

TABLE 46

PLAY ACTIVITIES ENGAGED IN BY MORE THAN 25% OF BOYS OF AGES 8½ TO 22 INCLUSIVE

Baseball with a hard ball. (November only.)	Listening to the radio.
Just playing catch.	Looking at the Sunday "funny" paper.
Riding in an auto.	Reading jokes or funny sayings
Watching athletic sports.	Reading the newspapers.
Going to the movies.	Reading short stories.
Chewing gum.	Reading books.
Card games, such as authors.	Writing letters.
Bridge, whist.	Whistling.
Listening to the victrola.	Teasing somebody.

TABLE 47

PLAY ACTIVITIES PARTICIPATED IN BY 25% OR MORE OF GIRLS OF ALL AGES FROM 8½ TO 22½ INCLUSIVE

Listening to the victrola.	Looking at the Sunday "funny" paper.
Playing the piano for fun.	Going to parties or picnics.
Riding in an auto.	Visiting or entertaining company.
Writing letters.	Chewing gum.
Reading short stories.	Teasing somebody.
Reading the newspapers.	Listening to stories.
Reading jokes or funny sayings.	Gathering flowers.
Going to the movies.	Just singing.
	Looking at pictures.

tial continuity of play interests. These are the play activities in which more than 25 per cent of boys and girls engage; *these activities are reasonably prominent in all ages from 8 to 22.*

These data say one thing clearly—that these common activities do not "belong" to any single age group. One would run into psychological error trying to place these activities in a planned program for one age but not for another.

Other tables appear in the original study which show the percentages of participation for each of the most frequently marked activities for each age. In getting away from the older notion that certain activities were natural ones at certain rather precise ages and not at other stages of development, we must also avoid an over-generalization in the opposite direction. We

cannot say that there are no tendencies for certain activities to be more prominent in the lives of boys and girls of certain loosely defined age groups. For example, "having dates" is very unimportant before ages 15 or 16, while "just running and romping" is more frequently checked by pre-adolescents. The point the investigators make, however, is that changes in play interests with age are very gradual and that there is considerable permanence of interests in a large number of favorite activities; changes are not sporadic and sudden, but are gradual and contingent on many variables.

Number of activities engaged in by individuals at different ages. Younger children not only engage in a greater variety of activities but also show greater child-to-child variability in number of activities. Figure 57 shows the data upon which this conclusion is based. The curves are plots of the median and the two quartile points for each age and sex. Eight-year-olds engage in about twice as many play activities as sixteen-year-olds. Their interquartile range is about twice as great indicating greater individual differences in number of activities. The shift is gradual; there are no sharply defined ages which one can designate as "age of all-around activity" or "age of narrowed interest," or any other "age." The trend of the curves is meaningful only if we stay clear of attempting to categorize ages.

Apparently as children grow older more conventional forms of entertainment govern play life while in children who are younger versatility of interests is more common. Also, it is possible that by age sixteen, duties in the home and work outside the home have increased their demands, thus limiting the variety of play opportunities in which adolescents engage simply through a limiting of time for recreation. It is also possible that the number of activities becomes narrowed as the children become older because they increase in ability to sustain attention on one type of activity. The number of activities is reduced, then, because these older boys and girls engage in each of their preferred activities for a longer time.

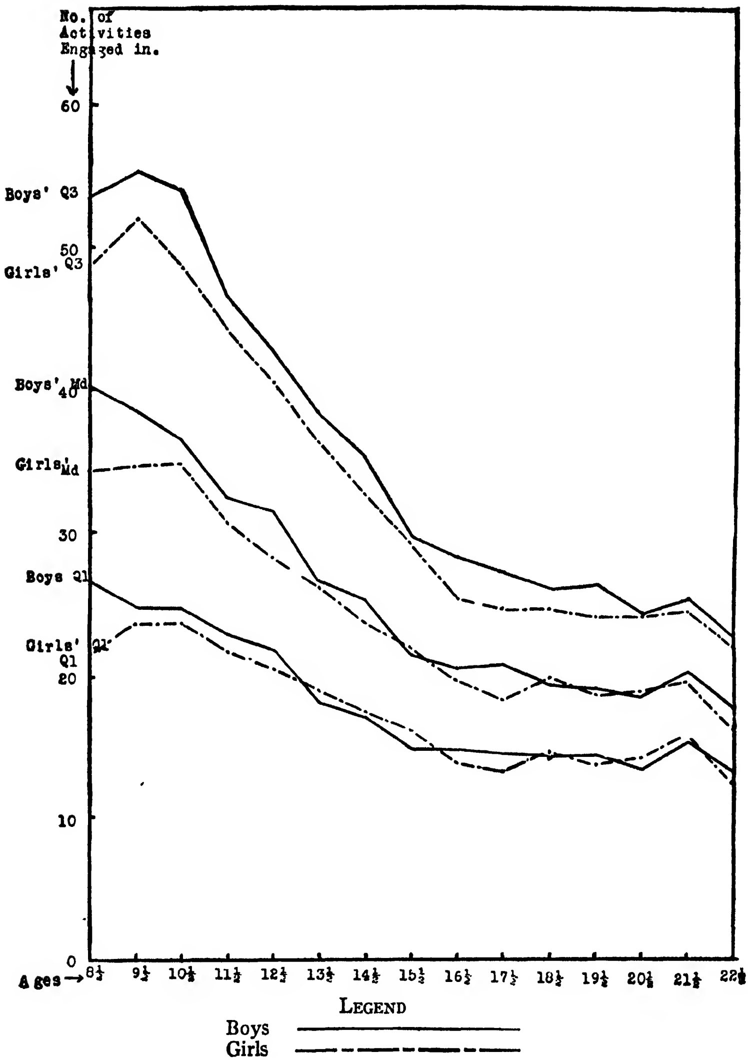


FIG. 57. The number of different play activities engaged in by boys and girls of various ages. Average of findings from three separate investigations. Dispersion of the two middle quartiles also shown. (From Lehman and Witty, *op. cit.*, p. 58.)

Transitoriness of play interests. While some games and play interests continue through the whole pre-adult span, there are some interests which are more transitory than others. Figure 58 and Table 48 present these data for boys. Similar data are available for girls. The length of each line represents the age span during which *at least* 25 per cent of the boys were engaged in the labeled activity. Fifty-five of the 200 activities are plotted.

This graph and the explanatory table need considerable study in order to force out of them the implications they can yield. Consider, for example, the meaning of item 48. Why do fewer than 25 per cent of boys between 10 and 18 visit and entertain company? Or, item 1, football: obviously the football games of eight-year-olds and of high school teams are not the same, but there must be some basic type of activity in common. Consider item 18; the roller skating "age" is brief. Are all of the activities below item 79 in the graph to be considered adolescent activities? How does this square with findings reported earlier?

Cautions are in order. (1) All of these activities are engaged in earlier and later than here indicated, these lines being only for the period during which *at least* 25 per cent of the boys reported participation. (2). The graph does not show peak periods of participation or interest. As a matter of fact, about two-thirds of the 200 activities for boys and girls showed modes in ages eight to ten. (3). Tremendous individual differences occur which cannot be plotted readily. (4). It is to be remembered that the findings here presented do not prove that classification of play programs on a basis of chronological ages is of no use; since physical structure, educational values, health considerations, and other similar matters also must be evaluated by the recreational planner. The data simply suggest that *artificial classifications may obscure a characteristic of all play behavior: its continuity from age to age.*

Social participation. Are there any characteristic ages during which children play predominantly alone or in groups? For children of school age and beyond, no such ages are found. For

TABLE 48

(KEY TO FIG. 58)

134	Playing in the sand.
187	Toy trains, ships, autos, wagons, etc.
183	Spinning tops.
48	Visiting or entertaining company.
77	Building or watching bonfires.
21	Coasting on a coaster.
32	Rolling an auto tire.
151	Playing Indian.
153	Playing robber and police.
24	Swinging.
105	Other tag games.
95	Follow your leader.
200	Playing with other pets.
99	Hide and seek.
150	Playing cowboy.
22	Coasting on a wagon.
175	Cutting paper things with scissors.
14	Checkers.
17	Marbles.
198	Playing with pet kittens.
67	Telling or guessing riddles.
88	Jumping for distance.
89	Jumping for height.
83	Just running and romping.
78	Climbing porches, trees, fences, posts, etc.
173	Drawing with pencil, pen, chalk, or crayon.
60	Playing the piano (for fun).
189	Looking at pictures.
86	Running races.
126	Throwing rocks or stones.
180	Using a hammer, saw, nails, etc., for fun.
29	Riding a bicycle.
197	Playing with pet dogs.
11	Boxing.
12	Wrestling.
1	Football. (November only)
133	Shooting a gun.
18	Roller skating.
2	Basketball.
79	Doing gymnasium work.
68	Telling stories.
165	Just singing.
28	Driving an auto.
196	Making or using a wireless or other electrical apparatus.
45	Going to parties or picnics.
44	Going to entertainments, concerts, etc.
51	Having "dates."
69	Listening to stories.
57	Social clubs, or being with the gang.
4	Ball with an indoor or playground ball.
52	Just loafing or lounging.
53	Social dancing.
50	Smoking.
42	Just hiking or strolling.
47	Attending lectures.

years 7 to 19, the percentages of *social* activities ranged from 63 per cent (age 8) to about 50 per cent (age 16). Plotting these percentages against age, the investigators got almost a straight line. In brief, for all of the school ages, about one-half, or a little more than one-half, of the play activities involved

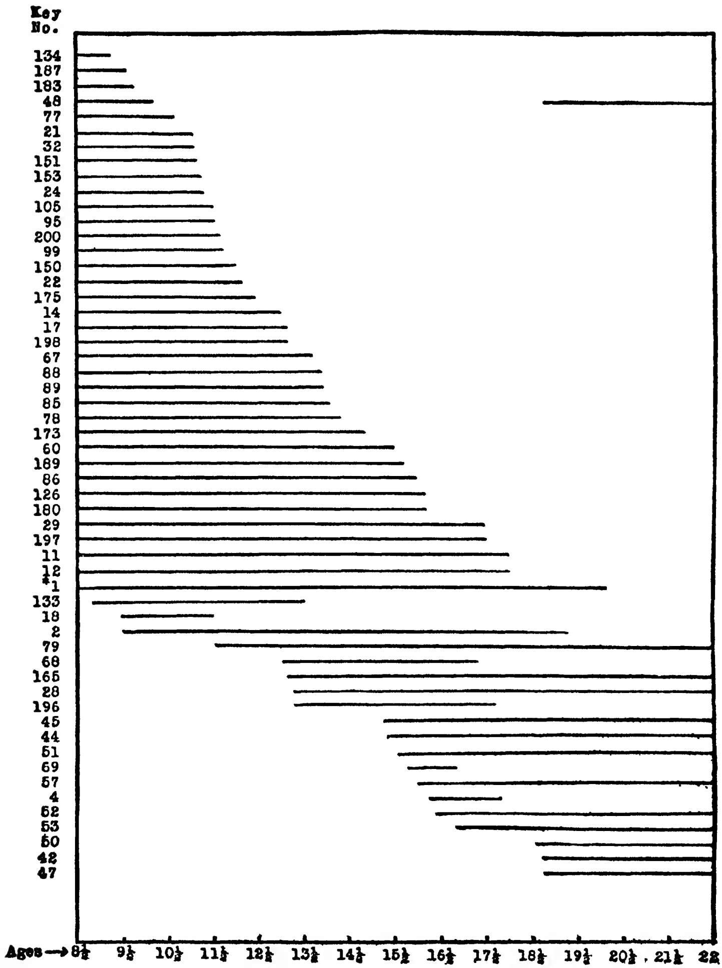


FIG. 58. Ages at which more than 25 per cent of boys were found to engage in various activities. Average findings from three separate investigations. (From Lehman and Witty, *op. cit.*, p. 64.)

more than one child; *i.e.*, they were social games and activities. Individual differences at each age were greater and more significant than age-group differences. There was a slight tendency for middle adolescents to be somewhat less social; but even then more than one-half of the play life involved at least one other person.

Types of play activities participated in by younger children as compared with older ones. Having pointed out the hazards of trying to classify children into "ages," the investigators nevertheless report that they can make a sketchy listing of the play activities by types of children 8 to 10 years of age and for upper ages. Their lower "age" is characterized by:

1. Activities involving pleasurable bodily movements usually of a rhythmic sort.
2. Activities involving hiding and finding.
3. Activities involving the imitation of adults.
4. Activities involving a relatively high degree of skill.
5. Activities which afford efforts at construction.
6. Activities which depend for their enjoyment primarily upon sense organ stimulation.
7. Tag games.
8. Singing games and ring games (for girls chiefly).

In making the list for older children, they merely eliminate items 2, 3, 7, and 8 from the above.

Discussion. We shall point out briefly some relations of these data, obtained by sampling the recreational choices of children by means of a check-list, to educational and vocational planning. These ideas are discussed, for the most part, in the final chapter of Lehman and Witty's book.

Supervision of play is fundamental. In addition to being a teacher of games and sports, of arts and crafts, and a "referee" in all sorts of social situations, the recreational supervisor is also responsible for identifying the solitary child and luring him into social games and for controlling the child who plays too much or in a too disorganized way.

Play Quiz data and data from other similar studies should

guide the curriculum makers. A curriculum should begin with the child as he is at the moment and provide a well-chosen program of activities which will help him progress along the educational line which is thought desirable. For example, if it is found that 50 per cent of the children in the seventh grade go to the movies regularly, say, once a week, that fact ought to be known by the schools. This information is pertinent to the seventh grade curriculum.

Expressed interests arise out of felt needs of the child. Since play occupies so much of the child's life, it follows that his play life will reveal many of his felt needs. Teachers, parents, and others who work with children need to observe their children's play life to diagnose the genuine needs of each child as they are there revealed. Children's insecurities and frustrations show up directly or symbolically in their free play. In psychological clinics it is quite customary to have observers report on the play life of the child in natural situations and in especially prepared set-ups under clinical auspices.

Play interests change from decade to decade. For example, listening to the radio was not a major feature of recreation before the mid-twenties; nor was the making of airplane models. Collecting and hoarding, Lehman and Witty report, were less frequently observed in the late twenties than they were in 1900. However, most of these specific modern activities have something in common with earlier interests. It has been said that all of our games are ancient.

Play interests have a bearing on vocational choices. Vocational choices depend, broadly, upon three factors: ability, opportunity, and interest. Play has direct connection with vocation in several ways. (1). Some vocations appear in elementary and playful form in childhood, as, for example, girls playing nurse or boys working with chemistry or skipping baseball to practice on the piano. (2). In play life one can observe the tendencies of children to prefer or not to prefer different kinds of social relations. The child who in spite of urging and opportunity still prefers to play alone (read, carve soap, or go on solitary nature-lore hikes) has revealed an important fact for

the vocational counselor as well as for those who look after the socialization process of school children. (3). Potential leadership may be discovered not so much from the way the child answers the Play Quiz as from observations of the child in social situations on the playground. Dominance and submission, crude and artful leadership, acceptance and rejection of cooperators—these are items to observe in play groups. (4). Play activities range from those which are almost entirely physical to those which are primarily intellectual or esthetic. No "all-or-none" boys and girls exist but some children lean more in one direction than the other in their interests and abilities. In both educational and vocational planning we want to take cognizance of a child's play preferences. (5). That play interests are important in measuring vocational interest is attested by the fact that Strong in his Vocational Interest Blank has included a section of recreational activities.[9] This instrument is based on the principle of measuring vocational interest by discovering patterns of "likes and dislikes" of successful persons in different vocational fields. Leisure-time interests seem to be important differentials in vocational guidance.

[9] Cf. Ch. 25.

Part Ten

SOCIAL PSYCHOLOGY

By

PAUL R. FARNSWORTH
Stanford University

Group Effects

Social effects. During this century a large body of material has been collected to demonstrate the effects of social stimuli on work output and general level of behavior. The famous Hawthorne study [1] has gone far to convince even the efficiency engineer that the factors which condition the amount of work a human being will do are not only the obvious ones of proper ventilation, lighting, rest periods, and pay. This research has shown that the attitudes of morale associated with the belief that the employer is attentive to the needs of his employees are closely knit to work output. But as yet the specific stimuli eliciting these attitudes have not been successfully isolated.

The studies of a laboratory sort have likewise been relatively unsuccessful in separating out the social variables one suspects are present. A few years ago Dashiell [2] classified the more experimental of the group studies according to the situations in which the subjects find themselves: where "(1) these group members may play the role merely of *passive audience* for the subject; (2) they may be *working alongside* him but not with any particular reference to him; (3) they may be contestants against him in that work; (4) they may verbally seek to affect his work with *remarks about him* or what he is doing; (5) they may cooperate with him by *interchange of ideas;* (6) they may be his sole source of information thus forcing him to rely upon *rumor* or hearsay; (7) in all these situations their influence upon him may be due to sheer numbers or *majority* or to personal *prestige.*" Each of the studies to be reported in this

[1] Mayo, E., *The Human Problems of an Industrial Civilization,* New York, Macmillan, 1933.

[2] Murchison, C., ed., *A Handbook of Social Psychology,* Worcester, Clark University Press, 1935, p. 1099.

chapter can, perhaps, be forced into one or several of these categories. However, as more and more data are gathered the complexity of these social stimuli and of the situations in which they arise becomes more obvious. As a consequence these seven categories must of necessity be gradually extended.

In the five sections of this chapter only a few samples of the many studies available can be discussed. And the conclusions from these must be taken as suggestive rather than as final. The social laws which have appeared are still quite indefinite and hardly worthy of such an impressive designation. Moreover, the complexity of the forces operating is such that many experiments, seemingly similar in conception and conduct, have yielded decidedly conflicting data. One can only hope that in time these inconsistencies will be reduced.

Social climates. The first main section of this chapter deals with an intriguing member of a series of researches that Lewin and his colleagues [3] have been making on the effects of several quite different social atmospheres. Our grandfathers were taught by the rod. We, perhaps, early learned that our state laws forbade even slaps from the teacher's hand. Are we, then, less frustrated by virtue of our gentle training? Or did our grandfathers accept their harsher atmosphere as "natural" and so rob it of most of its frustrating possibilities? While the Lewinian studies do not answer these questions, they at least open the field for further interesting research.

The J-curve hypothesis. Next comes a view of what is perhaps the first attempt to use quantitative methods in the study of institutional acts. Allport and Solomon [4] explain how they equate the ordinary mathematical continua of time and space with a novel type of continuum expressed in terms of the satisfaction of some goal. It is not enough, they believe, either to describe in purely qualitative terms or to express social phe-

[3] Lewin, K., Lippitt, R., and White, R. K., "Patterns of Aggressive Behavior in Experimentally Created 'Social Climates,'" *J. Soc. Psychol.*, 1939, *10*, 271-299.

[4] Allport, F. H., and Solomon, R. S., "Lengths of Conversations: A Conformity Situation Analyzed by the Telic Continuum and J-Curve Hypothesis," *J. Abnorm. (Soc.) Psychol.*, 1939, *34*, 419-464.

nomena in the measures of physical science—in centimeters, grams, and seconds. What we should like to know is something of another sort: Did the particular institution fulfill its function and to what degree? The J-curve studies are helping to answer these questions.

Social effects in physical isolation. The third study[5] carries the phenomena of social effects to a state of extreme dilution to prove that they are present even in the physical absence of other persons. As our very thoughts are socially conditioned we should not be surprised to find that when we know that we are attempting what others are currently doing elsewhere, there is a demonstrable social effect on our behavior. Moreover, the effect is in the direction forecast from other studies on more potent social stimuli, *i.e.*, there is more often an increase in the quantity of work done but a decrease in the quality.

Motor conflict. The fourth research[6] studies under strict laboratory conditions the types of behavior which follow artificially induced frustration. After carefully training a person to follow one sort of signal with a particular act and a second signal with an act antagonistic to the first, what happens when both signals occur at once? The arm cannot both be flexed and extended at the same instant nor can an army advance and retreat simultaneously. Nevertheless, in any culture man is occasionally called upon to perform what appear to him to be mutually incompatible acts. Frustration of some sort naturally results. Even animals have been made psychopathic by being placed in what to them must have seemed to be impossible situations. Having learned that food is given only when certain acts are performed, at a later feeding time they are forced to behave in a manner guaranteed to drive the food away. Such studies undoubtedly will contribute heavily to the abnormal and social psychology of the future.

[5] Dashiell, J. F., "An Experimental Analysis of Some Group Effects," *J. Abnorm.* (*Soc.*) *Psychol.*, 1930, *25*, 190-199.

[6] Hovland, C. I., and Sears, R. R., "Experiments on Motor Conflict. I. Types of Conflict and Their Modes of Resolution," *J. Exp. Psychol.*, 1938, *23*, 477-493.

Panic behavior. The last study shows what happens when our institutional habits go awry. It is said that army drill is mainly useful for times of stress. For, at a time when excitement might dull clear thinking, each man has an institutional act to perform. Similarly, the rules of etiquette keep us from colliding in a narrow doorway or eating our table-mate's salad while in the throes of social embarrassments. But when something unusual threatens and we have no institutional behavior adequate to the emergency, then panic ensues. To many Americans the "Invasion from Mars" [7] broadcast on October 30, 1938, was an occasion of this sort.

Social Climates

INTRODUCTION

Social laws are relative to social climates. Economists and social psychologists have been forced to accept the idea that their so-called "laws" function only within rather narrowly circumscribed boundaries. Totalitarian economics does not follow the same principles as does democratic-capitalistic economics. Specialists frequently forget this and wonder why their opponent's economic structure does not promptly collapse. Surprise has been expressed by other social scientists at the fact that financial ruin does not drive members of the "depressed" classes to suicide with at least the same relative frequency as it does members of the "monied" groups. However, this phenomenon should not prove puzzling as the two social classes obey somewhat different social psychological "laws"; their social climates are dissimilar. Two quite divergent habits of expectancy have developed, one to be surprised if any considerable success appears, and the other to anticipate an almost continuous sequence of successes.

[7] Cantril, H., *The Invasion from Mars: A Study in the Psychology of Panic,* Princeton, Princeton University Press, 1940.

ILLUSTRATIVE PROBLEM

An experiment on climates. A splendid illustration of the effects of social climates on behavior is shown in the recent study by Lewin, Lippitt and White. Clubs composed of 10-year-old boys were organized on a voluntary basis ostensibly to study mask making, mural painting, soap carving, model airplane construction, etc. Certain of the clubs were run on the "democratic" plan. These boys chose the activities they particularly desired. Other clubs, run on the "authoritarian" plan, had the activity decided by the leader. Still others, run on a "laissez-faire" basis, were told of the possible courses of action when problems arose, but were not otherwise influenced in their choices. An attempt was made to equate the several groups as to certain personality characteristics, congeniality, and intellectual, physical and socio-economic status. Every six weeks each club had a new leader and a new social climate, *i.e.,* "authoritarian" clubs became "democratic" or "laissez-faire," and "laissez-faire" became "democratic" or "authoritarian," etc. Two clubs met in adjacent but quite separate areas of the same large room. Observers sat behind a low burlap wall in a darkly shaded portion of the room from which both the leader and the boys could be observed.

Differences between climates. Table 49 describes the three atmospheres or climates. An "authoritarian" climate as rigorous as might have been desired quite obviously could not be established over a purely voluntary group of boys. As it was, several members left the groups, largely because of hostilities directed against them by other members of the "authoritarian" clubs.

Results. While the data tend to be more easily expressed in qualitative rather than in quantitative terms, certain numerical comparisons can be offered. Leader participation was less than half as great in the "laissez-faire" as in either of the other two groups. That the "authoritarian" leaders truly directed activity was shown in the observation that they did so sixty-three times as often as they complied with the sugges-

TABLE 49

Authoritarian	*Democratic*	*Laissez-faire*
1. All determination of policy by the leader.	1. All policies a matter of group discussion and decision, encouraged and assisted by the leader.	1. Complete freedom for group or individual decision, without any leader participation.
2. Techniques and activity steps dictated by the authority, one at a time, so that future steps were always uncertain to a large degree.	2. Activity perspective gained during first discussion period. General steps to group goal sketched, and where technical advice was needed the leader suggested two or three alternative procedures from which choice could be made.	2. Various materials supplied by the leader, who made it clear that he would supply information when asked. He took no other part in work discussions.
3. The leader usually dictated the particular work task and work companions of each member.	3. The members were free to work with whomever they chose, and the division of tasks was left up to the group.	3. Complete nonparticipation by leader.
4. The dominator was "personal" in his praise and criticism of the work of each member, but remained aloof from active group participation except when demonstrating. He was friendly or impersonal rather than openly hostile.	4. The leader was "objective" or "fact-minded" in his praise and criticism, and tried to be a regular group member in spirit without doing too much of the work.	4. Very infrequent comments on member activities unless questioned, and no attempt to participate or interfere with the course of events.

tions of the boys. The "democratic" leaders directed only 1.1 times as often as they complied. In seven cases out of ten the "laissez-faire" leader was preferred over the "authoritarian"; in nineteen out of twenty the preference was given to the "democratic" leader.

The aggressive actions per meeting were most frequent in the "laissez-faire" clubs, *38;* the value for the "democratic" was intermediate, *20;* the "authoritarian" clubs split into two categories with either numerous aggressive reactions, *30,* or apathy, *2* per meeting. That the apathetic boys were badly frustrated is shown by four lines of evidence. (1). When they went into freer atmospheres there were often sudden bursts

of aggression. (2). The apathy tended to disappear as soon as the leader left the room. (3). While the attitude of the apathetic boys was not that of extreme discontent, it was not by any means that of content. (4). Interviews with the boys disclosed an almost unanimous dislike of the "authoritarian" leaders regardless of the latters' particular personality traits.

Conclusions. The varied social atmospheres which surrounded several groups of 10-year-old boys were very effective in creating frustrated behavior. Among the ways in which frustration was displayed were aggression and a more disguised form, apathy. As the authors contend, aggression may be interpreted "in terms of four underlying factors: tension, restricted space of free movement, rigidity of group structure, and style of living (culture)" (p. 299). That tensions must develop under any sort of cultural system is obvious. That activities must be curtailed, space of free movement limited for the good of the greater number, cannot be denied. That certain groups are more rigidly constituted than are others should also be evident. But the fact that the style of living should be important is not always so easily recognized. If, for example, the above experiment could be performed in a totalitarian country where markedly different culture patterns prevail, quite different data might well be expected. Boys reared in an atmosphere where all initiative comes from above might react to the Lewinian "authoritative" climate quite differently. Care should be taken, therefore, not to carry conclusions too far afield or to apply them to a culture area which differs too markedly.

The J-Curve Hypothesis

INTRODUCTION

The classification of institutional acts. It has long been traditional for the social scientists to classify institutional acts as folkways, customs, and mores. Folkways are the "proper" ways of doing things within a given group. Customs are the "only" ways, while the mores make up the numerous tabus

and positive social commandments all races follow to a greater or lesser extent. But although this type of classification is useful, it is not in any sense quantitative. For this reason Floyd Allport, always on the alert for a chance to quantify, has introduced the concepts of the single and double J-curve distributions.

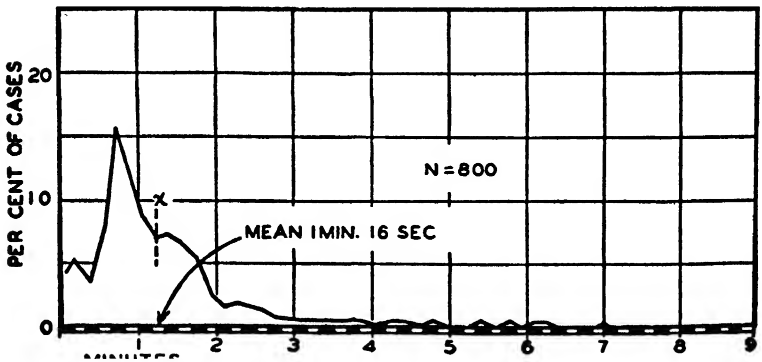


FIG. 59. Distribution of Duration of Conversations. Library situation. Empirical continuum (time).

AN ILLUSTRATIVE PROBLEM

The appearance of the double J. The double J is said to appear whenever highly institutionalized data are plotted in ordinary mathematical (empirical) terms. If, for example, one tabulates the duration of one common sort of library annoyance as Allport and Solomon have done with 800 whispered conversations at Syracuse University, there will emerge a double J-curve distribution. That is, the shape will resemble two sharp curves or J's placed back to back.[8]

The appearance of the single J. But what would happen if the same data were plotted in terms of some non-empirical or

[8] There has been considerable argument as to the extent to which the double J differs from the normal curve. See Dickens, M., and Solomon, R., "The J-Curve Hypothesis: Certain Aspects Clarified," *Sociom.*, 1938, *1*, 277-291; Dudycha, G. J., "The J-Curve Hypothesis: A Reply to Dickens and Solomon," *Sociom.*, 1939, *2*, 52-58.

telic continuum?[9] Library conversations can be plotted either in units of time or in terms of the satisfaction or blocking of some institutional aim. Libraries are not intended for conversations. If these latter are at all extended they cause other library users to be greatly annoyed, thus destroying, at least in part, the function of the library. It should be of great interest, therefore, to replot the conversation data in terms of annoyance and to see to what extent the institutional aspects of the situation cause a skewing of the cases to form a single J-shaped distribution. In carrying on such a study Allport and Solomon proceeded as follows:

Fifty-six students of elementary psychology at Syracuse University were asked to imagine themselves in a library situation in which two other people were conversing. Then the statements which appear in Table 50 were presented in ran-

TABLE 50 *

MEDIAN VALUES

Statements	*Degree of annoyance (inches)*	*Time of performance (minutes)*
A. Slightly distracted from what I am doing	.175	1.25
B. I begin to notice the conversation	1.25	1.58
C. Feel like stopping my activities and staring, or actually do so	3.65	3.15
D. Feel like getting up and asking them to stop, or actually do so	7.25	5.75
E. Feel like asking an authority to make them stop, or actually do so	8.85	7.90

* Adapted from Table 4 of the Allport and Solomon article.

dom order to the students who were to rate them in terms of the degree of annoyance they called forth. Each subject pasted the statements on a sheet of paper 9.5 inches wide. The statement representing the least annoyance was pasted somewhere at the left and those indicating more annoyance farther to the right. The distance from the left edge of the paper thus

[9] A telic continuum is one whose units are in terms of the satisfaction of a purpose or goal.

indicated the degree of annoyance implicit in the statement. The smallest degree possible was 0 and the greatest 9.5.[10] "The ratings were not performed in the classroom, but by the students working privately, so as to avoid any effects of social

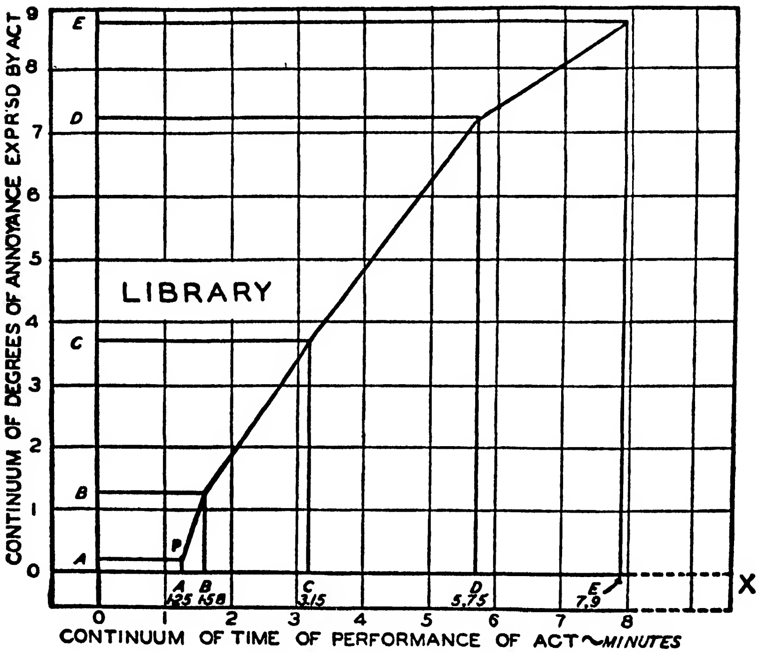


FIG. 60. Curve describing relation between degree of annoyance and time of performance of the five acts (library situation).

facilitation or conformity attitudes in the rating" (p. 437). The medians of the ratings are shown in the first column of numbers in Table 50.

The next step was to translate these degrees of annoyance into units of time. This was accomplished by presenting each rater a month or so later with a chart bearing 5 horizontal lines all slightly over 11 inches in length. "Each line was divided into uniform units by vertical divisions showing fif-

[10] For a somewhat similar sorting scheme see p. 522 ff.

teen-second intervals, with longer division indicating successive minutes up to eight, the number of the minutes being marked below. The rater's task was to check on each line his estimate of the time which the conversation would have to

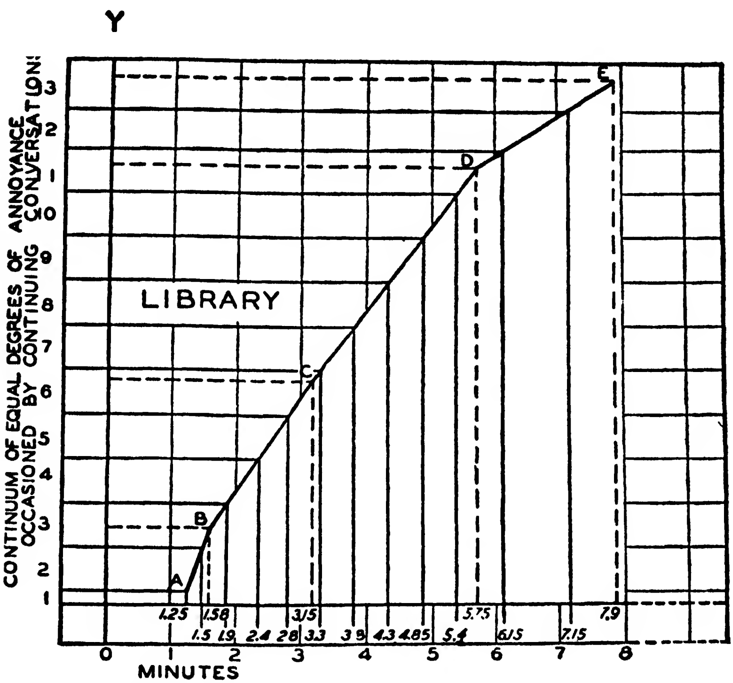


FIG. 61. Values of equal units of annoyance occasioned by conversation (Y-axis) expressed in terms of minutes during which the conversation lasts (X-axis) (library situation).

endure before the point would be reached at which he would do the act or experience the state in question. Should the time estimated run over eight minutes, he was asked to write it on the space provided over the right extreme of the line" (p. 439). He was to imagine a conversational tone a little above a whisper. The medians of these ratings are shown in the last column of Table 50.

The data of the two columns of Table 50 were then plotted as in Fig. 60 and the resulting curve was drawn.

The y axis was next rescaled as in Fig. 61 by laying off several equal intervals starting from the position of A. The area below A becomes Interval 1 which in this problem takes in 1.25 minutes (1′ 15″).

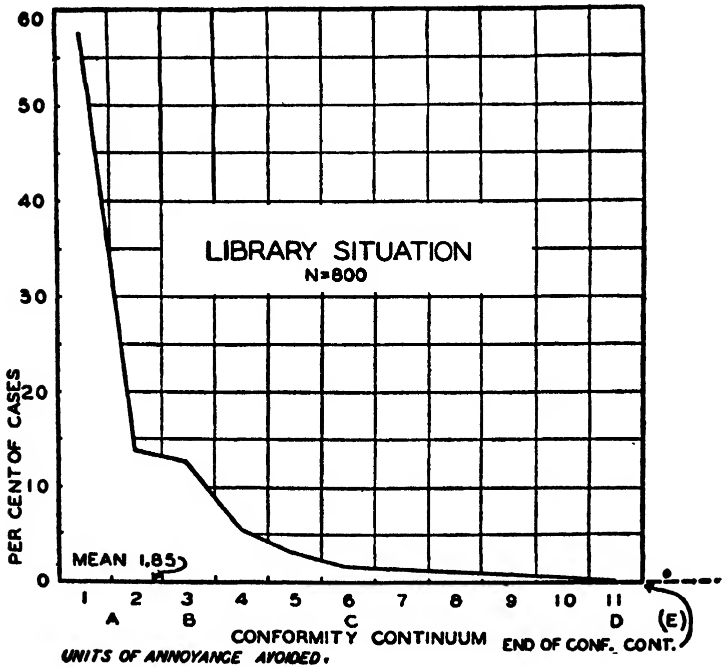


FIG. 62. Distributions of lengths of conversations in a library upon a continuum of conformity.

Its value—57.9 per cent—can be found from Fig. 59 by adding together the percentages at the plotting points up to the X mark (1.25 minutes). Further similar additions will disclose the values of the other intervals. From these data Fig. 62 can be constructed. Here we see a typical single J-curve.

Conclusion. Allport contends that a "field of conformity" exists whenever at least 50 per cent of the cases fall in the first interval of the continuum. The library situation, then, has met the criterion of conformity. Data presented in the Allport-

Solomon article but not offered here show that church conversations yield even more of a J-curve (84.0 per cent in Interval 1) while clubroom, fraternity and sorority conversational situations do not display such a curve of institutionalization (only 27.0 per cent in Interval 1). These differences are quite in keeping with common sense.

Somewhat similar but less complete analyses have been made of factory tardiness, of the behavior of automobile drivers at variously guarded street intersections, and of a number of other institutional phenomena. In all of these the J-curve appeared. But in other situations, where the institutional aspect was weaker, the distributions resembled the J far less. An interesting analysis of institutional behavior appears in a study [11] in which Baptists were queried concerning their beliefs on the matter of the proper mode of baptism. When consulted away from the church atmosphere there was much less insistence on the traditional Baptist mode of baptism. Two distributions thus could be made for Baptists, a clerical or institutional one with J-shape, and a non-clerical or less institutional one with a more "normal" shape. Of course no behavior is exactly 100 per cent institutional. If it were, there would be no continuum, as all persons would conform to one mode of behavior.

Social Effects in Physical Isolation

INTRODUCTION

Group effects ever present. It is almost axiomatic that man's behavior in a group differs from his behavior when alone. But it is not so generally recognized that he is always affected by his group whether others are physically present or not. The human baby could no more thrive without some sort of social medium than a tree without sunlight. Even Robinson Crusoe lived by making use of the bag of tricks he had learned

[11] Schanck, R. L., "A Study of a Community and Its Groups and Institutions Conceived of as Behaviors of Individuals," *Psychol. Monogr.*, 1932, *43*, No. 2.

as the result of his previous cultural training. Since group effects are present even for the solitary individual, the social psychological problem becomes that of learning their kind and extent. But the task is not an easy one. Out of the mass of experimental data so far gathered there has gradually emerged the realization that the effects are far more subtle and the variables much more difficult to separate than was formerly supposed. Even the question, "What is the effect on work output of working alone as compared with working in the presence of others?" is found to be at least a double one. That this is so can be seen from the fact that one can work (a) alone but with the knowledge that others are working at the same time and at similar work but in a different room, or (b) alone and at a time when it is certain that no one else is working at the same task. Quite obviously the former situation is more social than the latter, although it is social only to a relatively minute degree.

ILLUSTRATIVE PROBLEM

Dashiell's problem. Dashiell has studied the above problem, at least as it refers to fairly simple multiplication (two-place by two-place numbers), mixed relations (Pintner-Renshaw list), and serial word-association tests. In his research an attempt was made to separate the effects of the work situation on speed from those on accuracy, as it had previously been shown by several other experimenters that social stimuli usually affect these two measures of work differently. As only 16 college students served in this portion of the Dashiell experiments, the conclusions must be taken as suggestive only.

Directions and situations. In their formal directions the subjects were told to work "as *accurately* and as *fast* as you can." . . . "In each situation the test series included a 4-minute working interval on the multiplication test, a 4-minute interval on the mixed relations test, and three 1-minute intervals in close succession on the serial association test" (p. 193). In the situation termed *AS* in Table 51, the individuals worked

at the same hour but in separate rooms, their time signals being buzzers which were electrically controlled. In the *AD* variation, each subject came to the laboratory at a *different time,* that is, each worked at a time when no one else was working. He received his time signal from an automatic interval timer which he himself managed.

TABLE 51 *

Multiplication: Speed

	Higher
AD	4
AS	12

Multiplication: Accuracy

	Higher
AD	10
AS	6

Mixed Relations: Speed

	Higher
AD	4
AS	12

Mixed Relations: Accuracy

	Higher
AD	9
AS	6

Serial Associations: Speed

	Higher
AD	9
AS	7

* This table is adapted from Table 2 in Dashiell, *op. cit.*

Conclusions. The figures of Table 51 are in terms of the situation in which the subjects made their higher scores. Note that the students who were more subject to possible social motivations—the *AS*'s—gave their higher scores when speed was the measure. The *AD*'s, who were in about as non-social a situation as Dashiell could devise, yielded their higher scores when accuracy was the measure.[12] These data are in line with

the principle that social stimuli, if they affect work output at all, tend to improve speed and harm accuracy. The Dashiell data indicate that mere physical separation did not completely destroy the social aspects of the situation.

Motor Conflict

INTRODUCTION

Three types of motor conflict. A topic of considerable interest to present-day researchers in the fields of social psychology, psychiatry, and child study is that which deals with intrapersonal conflict or frustration. Lewin,[13] who has been primarily interested in the social psychology of children, has revitalized the work in this area with his novel experimental set-ups and his geometric designs of social forces. He has described three easily recognizable types of conflict which must be resolved in some way or other.

Type I, *approach-approach,* includes situations in which the organism is attracted to two objects and cannot immediately choose between them. As the two desires are not precisely equal in intensity, there is eventual resolution in the choice of one. Type II, *approach-avoidance,* supposedly underlies the ambivalent responses even relatively normal persons show toward their parents. For although our elders are traditionally our major source of comfort—in Lewin's terminology our parents possess positive valence—they are also forced to mete out most of the don'ts which loom so large during our early years, and to set up avoidance reactions toward themselves—in Lewinian terminology they thus possess negative valence as well. Our attitudes toward war may also be categorized as of Type II. We may be attracted by the idea of saving democracy, of travel, of attracting the ladies with our trim uniforms, of raising social status through the possession of military titles, or of

[12] The data of the serial associations test were not in line with those obtained from the other tests.

[13] Lewin, K., *A Dynamic Theory of Personality,* New York, McGraw-Hill, 1935.

being temporarily released from humdrum jobs. Yet, at the same time, we may be repelled by the thoughts of the terrible sights, sounds, odors, and above all the suffering that necessarily follow almost every war. Type III, *avoidance-avoidance,* occurs when we find ourselves urged to choose between two disliked activities. According to Lewin we resolve this dilemma most commonly through avoiding both activities by proceeding to other quarters.

ILLUSTRATIVE PROBLEM

A fourth type of motor conflict. In 1938 Hovland and Sears educed a Type IV conflict, a variant of II but somewhat more realistic, they felt. It embraces situations "in which the

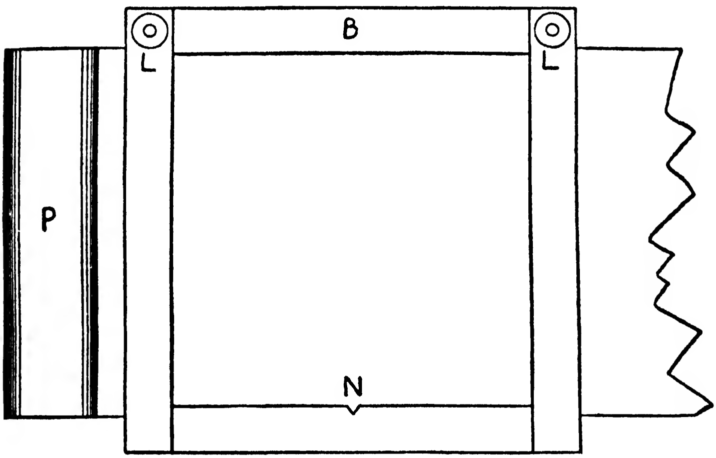


FIG. 63. Diagram of the conflict board. (P = roll of stylograph paper; B = brass border; N = nick in which stylus rests and from which reactions start; L = lights.)

organism faces two interlocking Type II situations at once (*e.g.*, a man has two desirable appointments at the same hour, the neglect of either of which will produce punishment or disappointment)" [14] (p. 477). To study this type and the other

[14] Sears, R. R., and Hovland, C. I., "Experiments on Motor Conflict. II. Determination of Mode of Resolution by Comparative Strengths of Conflicting Responses," *J. Exp. Psychol.*, 1941, *28,* 280-286.

varieties of motivational conflict the authors constructed an apparatus by which they could variously stimulate and frustrate their subjects and thus create situations similar to those mentioned as belonging to the four types. For the purposes of the present report we shall describe their methods of creating motor conflicts of Types II and IV only and the resolutions made by their subjects.

Apparatus. The subject sat at a table on which was a smooth metal plate, 6 x 6 inches, which was bordered by brass strips 1 inch wide and ¼ inch thick (B). The two strips were raised ¼ inch so that 6-inch-wide stylograph paper (P) would slide through and serve as a wax surface 6 inches square on which the subject could mark. At the middle of the side toward the subject a small nick (N) was cut in the inner edge of the brass strip. The subject rested his pencil in this at the beginning of each trial. . . . For a study of Type II conflicts a pair of sockets [was] arranged diagonally so that each member of the pair was approximately equi-distant from the nick in which the pencil rested. A red and a green flash-light bulb were placed on one corner. For Type IV conflicts red and green pairs of lights were placed at both corners (p. 478).

Instructions and procedure. Group II. A red and green light were placed at one corner of the board only and the subject was instructed to go *toward* the flash of one (*e.g.*, green) and *away* (opposite corner) from the flash of the other. On the 21st trial the two lights were flashed simultaneously (approach-avoidance conflict) (p. 479).

Group IV. Pairs of red and green bulbs were placed at both corners of the board and the subject was instructed to go *toward* the green light and *away* (opposite corner) from the red light, no matter on which side a light appeared. Since this situation involved differential reactions to 4 lights instead of only 2, more trials to each light were given before the conflict situation was presented (p. 480). The conflict was caused by the simultaneous flashing of all 4 lights.

Response types. Four modes of response were noted.

(a) *S—Single responses.* When a subject resolved his conflict in what the authors term an "S" manner, he drew a line directly to one of the two lights.

(b) *D—Double reactions.* These usually involved going toward one signal, returning to the nick, and then proceeding to the opposite corner. Occasionally the subject drew a line to one light and then went directly across the top of the board to the other light.

(c) *C—Compromise movements.* A compromise involved going up the center of the board midway between the two corners.

(d) *B—Blocking with no manual response.* The authors suggest that blocking is not a permanent resolution and may represent merely an unusually long reaction time.

TABLE 52 *

PERCENTAGES FALLING IN EACH MODE OF RESOLUTION

Conflict type	*S*	*D*	*C*	*B*
II	31.8	13.6	0	54.5
IV	10.0	12.5	5.0	72.5

* This is taken from Table 2a in the Hovland and Sears article.

Conclusions. The most striking characteristic of these two types, particularly that of IV, is the large percentage of blockings. (Data presented in the original article show that blocking is also quite characteristic of Type III with double reactions as least representative. Type I is characterized by more single responses.) Hovland and Sears admit that these somewhat artificial situations may be a far cry from major emotional conflicts. Yet the findings are probably somewhat indicative of what takes place outside the laboratory. And at least they furnish a conceptual framework which should aid one in describing the several varieties of frustration.

PANIC BEHAVIOR

INTRODUCTION

Panic behavior little observed. The behavior of people in panics has been but slightly understood because, with little or no advance warning, social scientists are rarely on the scene

to record the events. Recently, however, there occurred an indisputable radio-induced panic at a time when a social psychologist, Hadley Cantril, was studying the psychology of radio audiences of his neighborhood. The discussion below grows out of the data Cantril and his aids presented in his book, *The Invasion from Mars.*

AN ILLUSTRATIVE PROBLEM

The broadcast. At 8 P.M., E.S.T., October 30, 1938, Orson Welles and his troupe broadcast an adaptation of H. G. Wells' *War of the Worlds.* At the outset an announcement was clearly made to the effect that what was to follow was only a play. But, unfortunately, many people tuned in late; the matter was complicated by the fact that a more popular program was scheduled at the same time on another station and large numbers of listeners tuned off this only to tune in on the Welles program at one of its most thrilling points. The continuity of the latter broadcast called for periodic "news bulletins" which were admittedly realistic. In brief they described astronomical disturbances in Mars which were followed in a short time by earthquakes centering near Princeton, N. J. The quakes were soon found to have been caused by several unusual-looking meteorites out of which crawled terrifying creatures equipped with tools of warfare far more deadly than any so far devised by human engineers. Battles began which spread as far north as New York City and then broke out in other areas in America as more meteorites landed.

Number and character of radio audience. From a pooling of the data collected in sample polls of the radio audience by the American Institute of Public Opinion and C. E. Hooper, Inc., it seems fairly certain that at least six million people listened to the broadcast. Roughly one million, seven hundred thousand mistook the play for a series of news flashes and about one million, two hundred thousand were disturbed by it. From a consideration of one sample of listeners reached by the Columbia Broadcasting System it would seem that the more educated the subject the less apt he was to think the play

real news. This trend held even when economic status was held constant.

The mistaken subjects were divided into four groups. There were first those who were said to have the greatest "critical ability." These analyzed the internal evidence of the program and knew that the "news" could not possibly be true, e.g., recognized Welles' voice, found time sequences impossible, etc. The members of the second group checked up successfully to learn that the supposed news was not real, e.g., tuned into another station. Those of the third group checked up unsuccessfully and so continued to believe in the validity of the "news," e.g., looked out of the window and saw an unusual glow. The members of the last group made no attempt to check the broadcast. In general those who failed to make adequate checks were the ones frightened. While data from several samples were not in perfect agreement, it can safely be said that not over a quarter of the mistaken subjects fell in the first group, while from a third to a half attempted no check.

Those who failed successfully to check the validity of the broadcast—the too suggestible subjects—behaved as they did for a number of reasons. First, they may have "possessed standards of judgment that adequately accounted for the events and made them consistent with latest expectancies," e.g., certain religious subjects believed these happenings to be in harmony with the predicted final hours of the world. Second, they may not have had "adequate standards of judgment to distinguish between a reliable and an unreliable source of confirmation," e.g., although other stations were tuned in, their "absence of news" was interpreted as an attempt on the part of the broadcasters to calm the populace. Third, they may have "had no standard of judgment and [may have] felt the need of one by means of which they could interpret the reports, thus accepting the interpretations provided by the 'observer' of the events and by the prestige of the radio." Lastly, they may have "had no standard of judgment and [so] unhesitatingly accepted the one provided" (p. 197).[15]

[15] Compare the findings with the material in Chap. 2, "The Frame of Reference"

The listening situation varied from person to person. Not only did the broadcast elicit powerful stimuli but telephone calls from excited friends, the arrival during the broadcast of hysterical relatives prepared to flee they knew not where, and yet other potent stimuli affected many listeners. Such possible complications cannot be neglected in a study of panics.

Detailed interviews were arranged with about a hundred radio listeners. Data gathered in previous Princeton radio studies had made it appear that listeners who typically felt insecure, whose worries were frequent and too intense, who lacked confidence in themselves, who had fatalistic attitudes toward the world and who showed by unusually strict church attendance or extreme observance of religious forms a degree of fanaticism, would prove to be especially vulnerable to radio suggestion, would be characteristically susceptible "to-suggestion-when-facing-a-dangerous-situation" (p. 130). For these reasons very crude ratings were given in each of these areas of weakness: "+" if the subject definitely fell into the category, *i.e.*, felt insecure, worried, etc.; "—" if he did not; and zero if the matter was in doubt. As might have been expected, those who had made successful checks of the validity of the "news" broadcast showed fewer of these weaknesses (scored 50+'s), than did those who made unsuccessful orientations (129+'s).

Conclusions. Panics can be said to occur "when some highly cherished, rather commonly accepted value is threatened and when no certain elimination of the threat is in sight. The individual feels that he will be ruined, physically, financially or socially" (p. 199). People who avoid succumbing to panics are those, more often than not with education, who are able to understand the inherent characteristics of a situation so that they can judge and act appropriately; in other words, they have critical ability. They are not very susceptible to suggestion when facing a dangerous situation or to emotions generated by an unusual listening situation.

CHAPTER 35

Interests and Attitudes

Questioning still the best approach. Although the study of brain waves may hold great promise for the psychology of the future, neither it nor any other physiological technique which studies the nervous or reaction systems can tell us much about interests and attitudes. Even this late in the twentieth century these latter are disclosed most adequately by some form of questioning. True enough, the questioning may not always tap basic interests and attitudes. We cover our true thoughts, we rationalize our deeds to convince ourselves, as well as others, that our actions are logical. Quite often our words do not fit the behavior which follows. Yet, for all of these difficulties, we must try to study the verbal side of life, for language is our most conventionalized form of behavior, the sort which differentiates us most clearly from the subhuman animals. In the sections which follow, samples are offered of only a few of the many methods of questioning. Notably lacking, for example, is mention of the personality inventory, now familiar to many students as a set of standard questions designed to elicit information concerning worries, angers, compulsions, and the like (this tool is discussed in Chapter 37). Lack of space has also caused other important omissions. Note should be made of the fact that the several methods of questioning overlap and that two or more are often used together.

The case history. The first paper below opens with a discussion of the least scientific but perhaps the most vital of the approaches, the case or life history method. Although Dollard[1] and others have attempted a standardization of the method, it still covers almost any sort of questioning whose

[1] Dollard, J., *Criteria for the Life History*, New Haven, Yale University Press, 1935.

intent is to bring to the surface attitudes and interests of a non-social and even of an anti-social type. These attitudes must be discussed with the subject in an attempt to integrate them into a more acceptable mental hygiene framework or life pattern. The success of the method depends in large measure on the skill of the operator. It cannot be employed blindly.

Interests and attitudes formally tested. The second study describes a few of the more highly statistical approaches traditionally employed in the studies of interests [2] and attitudes. Although the more theoretically minded of the social psychologists claim that one should differentiate between these terms, in the practical testing field they have become almost interchangeable. In the attempts to study them, brief statements or queries are commonly presented to the subject with the request to check if he agrees. At other times he may be asked whether he likes, dislikes or is indifferent to a particular item. On occasion the items appear in pairs to be evaluated (see p. 544), or must be ranked in order from most to least liked, or given some numerical value. Sometimes the subject is offered several alternatives in the solution of some social problem; for example, "If a Negro were seated at your restaurant table would you (a) proceed with your meal? (b) ask that he be given a different table? (c) leave the restaurant?" etc. Luckily these rather different-appearing methods tend to yield essentially similar pictures, at least as far as group results are concerned. Hence the choice of a method becomes more or less a matter of convenience and taste.

Public opinion polls. The interview method as expressed in the selection of sample cross-sections of the buying public or of the electorate has become of use to business and politics through the joint efforts of political scientists, journalists, and social and business psychologists. The utter collapse of the *Literary Digest* polling service has made it strikingly clear that most careful attention must be paid to possible sampling errors in the selection of straw voters as well as to the precise

[2] A discussion of the Strong Vocational Interest Test will be found in Ch. 25.

phraseologies of the questions asked. Each modern polling service has its own set of weights to be used in the selection of just those subjects who will be properly representative of the later electorate or of subsequent buyers. The problems still bothering the commercial poller are legion; however, the methods are already so good that to say that straw votes are often better indices of real attitudes than are actual ballots is perhaps only a slight exaggeration.

Level of aspiration. The last study reported in this chapter is concerned with a phase of attitude measurement rather recently opened to experimental investigation. It deals with the relations which exist between a person's attitudes toward his own successes and failures and the degree of achievement which follows. A consideration of the few pioneering studies so far in print makes it clear that the social factors which operate are quite complex. The level of aspiration is quite unmistakably a function of a man's intelligence, his basic habits, interests and attitudes, and his momentary state which in turn is in part a function of the particular constellation of stimuli operating at that moment. The Gardner study [8] is featured in this chapter as it controls fairly well at least one variable, the behavior level believed to have been just achieved. Other variables need to be somewhat similarly isolated.

THE CASE METHOD

INTRODUCTION

The case method an art. The case method appears in many forms. Each clinician has his own highly prized mode of attack through which he hopes better to understand his patient. His rules and those of his fellow clinicians make up the art which underlies the case method. But in no true sense can these principles be said to form a science. For even with the best of present-day techniques, the subject may withhold the most pertinent facts about himself. Moreover, it can never be proved

[8] Gardner, J. W., "Level of Aspiration in Response to a Prearranged Sequence of Scores," *J. Exp. Psychol.*, 1939, *25*, 601-621.

beyond reason of doubt that the correct interpretation of the available data has been made in any specific case. For these reasons many rather bizarre techniques and interpretative theories have crept into the literature of the case or life history method. Many times a psychoanalytic slant of some sort is given; but more typically [4] the procedure includes interviews, both formal and informal tests, ratings, and the writing of short autobiographies. Contributory information is often furnished by the social worker, school teacher, physician, and by others.

AN ILLUSTRATIVE PROBLEM

Description of symptoms. In the 1933 files of a certain clinician is the life history of a college lad, C. I. M., who came in for an interview because of rather specific adjustment difficulties. As his situation was not extremely serious the interview alone was employed. Records were already at hand as to health, college grades and Binet intelligence. These showed a healthy boy of average intelligence—an IQ of 105—but a lad with far more than average success in the obtaining of grades. Such a combination suggested a "grind" and indeed this was one of the brands bestowed upon the lad by his associates. He was also called snobbish, disdainful, uninteresting, and was known to be outspokenly anti-Semitic. He admitted the validity of these descriptions but claimed not to know how he had come to possess these characteristics. He desperately wanted friends. He willingly came for weekly interviews for slightly more than an academic quarter. During this period the following story appeared bit by bit.

Background data. C was born of German parents in a small Middle-Western community. His father, a husky lumberman, had the continental attitude toward discipline. His mother, a

[4] Chassell, J. O., *The Experience Variables Record,* Rochester, N. Y., 1928, has been widely used in case history work. See also Stogdill, E. L., and Herndon, A., *Objective Personality Study,* New York, Longman's, 1939. For an extensive study which made use of many procedures see Murray, H. A., *et al., Explorations in Personality,* New York, Oxford University Press, 1935.

scholarly type of woman who had clearly married beneath her mental and social station, pampered the youth and took his side in all father-son quarrels. Soon C idolized her and grew to have her interests. He avoided as much as possible the father who quite naturally regarded him as a sissy. The father's attitude waxed more and more hostile as it became apparent that the boy would take no part in rough games or "manly" sports.

Prodding on the part of the clinician disclosed the fact that the boy had early found that he was not brilliant. But his desire to achieve superiority in intellectual fields and the excellent study habits developed through the aid of his mother gave him high rank in school. C gladly spent long hours on his books as a sort of overcompensation for his lack in rough-and-tumble games. Physical competition of all kinds came to be associated first with his father and then with the rougher boys of his neighborhood. All such behavior was boorish and to be looked upon with high disdain. All things intellectual, particularly those which he had mastered, became highly important.

C's anti-Semitism could seemingly be traced to an unfortunate situation in 1918. When the United States entered the world war, C was five years of age. At that time he spoke only German and was steeped in the lore of the "fatherland." During the period in which harsh feelings were mounting against all things German, C was jeered at, his house was pelted with over-ripe fruit and he himself was badly mauled on several occasions. As his tormentors were all his physical superiors, he at first saw no way to retaliate. But later he noted that several of the lads were spoken of as Jews. As soon as he learned that this label showed them to be somewhat vulnerable he made the most of it. He even combed the literature for ammunition he might employ against the Jews. And the fact that his closest competitors in school were Jewish did not add to his love for members of their "race."

The clinician's interpretation. The clinician took the not unreasonable stand that C's lack of friends was not associated with his being a grind but rather with more fundamental fac-

tors. He was told that he probably had built up his present much-disliked constellation of attitudes as a protection against his father's scorn and his own felt inferiorities, physical, mental, and social. He had seized upon study and anti-Semitism as overcompensation for these inferiority feelings and had assumed a snobbish exterior as a cover for his wounded feelings. C agreed that he no longer had reason to fear either social or physical inferiority. In 1933 he no longer was being branded as a "Hun," and he had reached the age at which prowess in physical combat was no longer the measure of a man. And, after considerable argument, he admitted that his scholastic achievements had proved that his rather ordinary IQ was not serving as a serious handicap.

Aftermath. Through these sympathetic interviews the boy was finally enabled to reverbalize his troubles into more sensible terms. During the months which followed, C made reasonably successful attempts to broaden his interests and to appear less aloof and disdainful. Although his grades suffered a little, they were still fairly good. While still disliking what he thought were Jewish characteristics, C even was able, after a short time, to accept a Jewish boy as an intimate. And, above all, the lad was happier than he had been at any other period of his life.

Formal Attitude Testing

INTRODUCTION

The Thurstone technique. Quantitative methods of a rigorous sort were brought to the field of attitude[5] measurement by Thurstone.[6] The technique he proposed, borrowed from the psychophysical method of equal-appearing intervals, is a simple one. Many statements having to do with a single aspect of the attitude to be studied are gathered. A table with some odd number of divisions—say 11—is arranged so that intelligent

[5] Sociologists commonly prefer the term "opinion."

[6] Thurstone, L. L., and Chave, E. J., *The Measurement of Attitude*, Chicago, University of Chicago Press, 1929.

judges may sort the statements into piles. Thus, if the attitude to be measured refers to war, pile 11 may arbitrarily be taken as the most militaristic, pile 6 as neutral, and 1 as the most pacifistic. The judge or sorter places each statement into the pile which, in his opinion, best represents the degree of militarism-pacifism expressed by the item. From the original list of items the experimenter seeks to find approximately two items within each step of the scale for each form of the test.[7] Every item must meet the following criteria: (1) it must receive a fairly consistent scale value from all of the judges; (2) it must be endorsed by enough people to make it a useful part of the scale; and (3) it must *not* be frequently endorsed by those who also check items having markedly different scale values. The median value of the sortings gives each item its scale value.

The unambiguous items are put into test form and presented to the groups whose attitudes are to be measured. Each person taking the test checks the items with which he agrees. The median of the scale values of these items becomes his score. A degree of validation is assured whenever staunch militarists and pacifists check only the items at opposite ends of the scale. If, by any chance, both groups should consistently check a given item (see criterion 3 above), it would be discarded as of dubious worth.

OTHER TECHNIQUES

Seashore-Hevner technique. As the technique of sorting is laborious and can usually be done by only one person at a time, modifications of the Thurstone procedure of finding item weights have been suggested. In 1933 Seashore and Hevner[8]

[7] A simpler 5-step scaling technique developed by Likert, R. ("A Technique for the Measurements of Attitudes," *Arch. Psychol.*, 1932, No. 140), is now coming into general use. Its chief defect seems to be that since it employs no judges there is no guarantee that Likert scale items will represent all degrees of the attitude to be measured. See Ferguson, L. W., "A Study of the Likert Technique of Attitude Scale Construction," *J. Soc. Psychol.*, 1941, *13*, 51-57.

[8] Seashore, R. H., and Hevner, K. A., "A Time Saving Device for the Construction of Attitude Scales," *J. Soc. Psychol*, 1933, *4*, 366-372.

offered a technique which placed all the scale statements of a given test on a single set of sheets. Beside each statement were the numbers 1 to 11 inclusive with the order to encircle the appropriate number. Seashore and Hevner found that their method yielded results which agreed closely with those obtained by the traditional method.

Ballin-Farnsworth technique. More recently Ballin and Farnsworth [9] have suggested the use of a sorting line. The judge places dots at the appropriate places along the line with the item numbers over the dots (see Fig. 64). Usually a sample placement dot is given to aid in the judging. The scale value is the distance from the left end of the line to the dot.

TABLE 53 *

MEDIAN SCALE VALUES

Statement No.	*Thurstone*	*Seashore-Hevner* †	*Ballin-Farnsworth*
4	0.2	0.6	0.4
7	0.8	1.5	1.1
9	1.4	1.0	1.1
12	2.1	1.7	1.7
17	2.4	2.6	2.9
19	3.2	2.2	2.6
2	3.5	2.2	3.0
14	3.7	1.7	2.6
11	4.7	4.7	4.8
8	5.5	5.4	5.3
16	6.5	6.6	6.2
13	6.8	7.1	6.5
5	6.9	5.3	5.5
1	7.5	7.4	6.6
10	8.3	9.2	8.9
6	8.7	8.4	8.3
20	9.2	9.9	9.7
3	9.7	10.1	10.0
18	10.1	10.0	10.0
15	11.0	10.4	10.7

* Adapted from Table 1 in the Ballin and Farnsworth article.
† Each of the numbers (1-11) was considered as the top of its class interval.

[9] Ballin, M. R., and Farnsworth, P. R., "A Graphic Rating Method for Determining the Scale Values of Statements in Measuring Social Attitudes," *J. Soc. Psychol.*, 1941, *13*, 323-327.

Comparison of the three techniques. In the Ballin-Farnsworth illustration the Thurstone-Peterson Scale of Attitude Toward War, Form A [10] was employed. Groups of undergraduate students of sociology and psychology, ranging in size from 100 to 200, served as judges. Table 53 shows the item weights as judged under each of the three techniques—Thurstone, Seashore-Hevner, and Ballin-Farnsworth. Note their essential similarity—which is also shown in Table 54 in terms of correlations.

TABLE 54 *

INTERCORRELATIONS

	Seashore-Hevner	*Ballin-Farnsworth*
Thurstone	.970	.984
Seashore-Hevner		.994

* Adapted from Table 2 in the Ballin and Farnsworth article.

By employing any one of these three methods, then, it is possible to find items with agreed-upon weights. Such items can be employed in the construction of attitude scales.

PUBLIC OPINION POLLS I

THE QUESTION FORM

INTRODUCTION

Speech vs. action. One of the most important problems of social psychology involves the learning of the degree of relationship which exists between the doing of a certain act and the verbal statement that it will be done. That is, with what accuracy can it be predicted that a man will do what he says he will do? From the data so far gathered it is clear that the answer depends upon the situation. In an area of behavior such as that of morals, it has been found that verbal tests give little or no help in forecasting subsequent behavior. In the field of esthetics, verbal responses have at least a little prognostic

[10] Thurstone, L. L., and Peterson, R., *Scale of Attitude Toward War*, Chicago, University of Chicago Press, 1931.

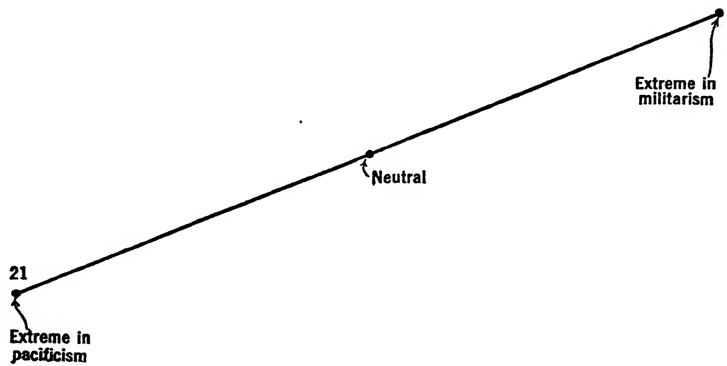


Fig. 64

On the appended sheet are 21 statements about war. You are to read them carefully and decide how pacifistic or militaristic they are. Note that the extreme left of the rating line is the pacifistic end and the extreme right the militaristic. Item number 21 is already rated for you, i.e., its placement on the line represents the judgment of many previous raters. It was regarded as completely pacifistic in character. Rate the other 20 items similarly. Put dots on the line at the points which you feel best express the degree of pacifism-militarism of the item. Place the numbers of the items over the dots. More than one item may be placed on the same point on the line.

THURSTONE-PETERSON SCALE OF ATTITUDE TOWARD WAR

Form A

1. Under some conditions, war is necessary to maintain justice.
2. The benefits of war rarely pay for its losses even for the victor.
3. War brings out the best qualities in men.
4. There is no conceivable justification for war.
5. War has some benefits; but it's a big price to pay for them.
6. War is often the only means of preserving national honor.
7. War is a ghastly mess.
8. I never think about war and it doesn't interest me.
9. War is a futile struggle resulting in self-destruction.
10. The desirable results of war have not received the attention they deserve.
11. Pacifists have the right attitude, but some pacifists go too far.
12. The evils of war are greater than any possible benefits.
13. Although war is terrible it has some value.
14. International disputes should be settled without war.
15. War is glorious.
16. Defensive war is justified but other wars are not.
17. War breeds disrespect for human life.
18. There can be no progress without war.
19. It is good judgment to sacrifice certain rights in order to prevent war.
20. War is the only way to right tremendous wrongs.
21.* Every last one of us should refuse to take part in any way in any war.

* Item 21 is from Form B.

value. But it is in the political field, where opinion polls have recently come into common use, that the connection between the verbal and the subsequent action is close. Presumably this is because of the similarity of the behavior in the two situations. In the Blankenship experiment [11] discussed below the significance of the form of the interview question has been studied.

ILLUSTRATIVE PROBLEM

The several question forms. In June, 1939, Blankenship was anxious to forecast the results of the New Jersey election on betting and horse racing. In constructing the interview questions he was forced to decide between a number of possible phrasings. He could have put the question into *objective* form. That is, he could have employed both positive and negative phrasings but not the word "you." ("Is it desirable to permit or to prohibit horse racing and pari-mutuel betting in New Jersey?") An alternate, the *subjective* form, contains both positive and negative phrasings and the word "you" as well. ("Would you vote for or against the amendment to permit horse racing and pari-mutuel betting in New Jersey?") The *positive objective* form is a modified version of the *objective* in which the negative phrasing is eliminated ("or to prohibit" is deleted). The *positive objective with check-list* contains the possible answers "Yes, No, or Don't Know." The *negative objective* form is a modified version of the objective in which the positive phrasing is eliminated ("to permit" is deleted).

Description of subjects. Preliminary work on the forms of the questions had been attempted on 100 subjects. For the major portion of the experiment, door-to-door interviews were taken on 3000 citizens of Irvington, N. J., from one to ten days before the election. The subjects were divided into five groups and each was given a different form of the question. The groups had previously been reasonably well equated for race,

[11] Blankenship, A. B., "A Correction," *J. Appl. Psychol.*, 1941, *25*, 124-127; "The Influence of the Question Form upon the Responses in a Public Opinion Poll," *Psychol. Rec.*, 1940, *3*, No. 23, 349-422.

sex, age, country of birth, religion, economic status, type of house, intention to vote, and ownership of automobile, automatic refrigerator, and telephone.[12]

Validities. The validity of each of the five phrasings was determined by comparison with the results of the election returns in which 50 per cent of the Irvington votes were cast for the horse-racing bill and an equal number against it. The ballots of those who answered the *objective* form of the question agreed most closely with the actual returns, with 49.5 per cent favorable. The *subjective* phrasing appeared to be the least valid, with 57.5 per cent voting favorably. The *objective* phrasing's closest competitor was the *negative objective* which gave 53.6 per cent of favorable votes. To quote Blankenship:

> The objective and negative objective question forms both predict the election returns within a chance basis; the other forms predict with such a degree of inaccuracy that this deviation cannot be accounted for by chance. Further analysis shows that these two forms were not reliably different from one another in terms of the affirmative (or negative) answers secured (p. 3).

Other calculations—not described in this book—showed the *objective* form to have least suggestion implicit in it. Yet this form, which had won on the above two counts, received the poorest ratings on consistency (stability in situations in which the same question is asked of an identical group on two occasions) and on indecision (frequency of "don't know" responses). However, these defects were not considered serious as the four phraseologies did not differ greatly in the matters of consistency and indecision.

Conclusion. So far as we may judge from this fairly extensive trial the *objective* forms, then, appear to be the best of the phraseologies so far studied. The major polling services are rapidly becoming aware of the importance of this type of study. It is now quite common practice to put questions

[12] The political polls usually weight their samples in terms of such items as state-federal population ratio, income, farm-city ratio, party, age and sex ratio.

through a variety of preliminary trials to make reasonably certain that the best possible question forms are being employed.

Public Opinion Polls II

The Band-Wagon Effect

INTRODUCTION

Band-wagon effect proposed. Does our knowledge that the betting odds favor a certain candidate make us more likely to vote for him? If the answer is "Yes," we are said to be responding, at least in part, to the band-wagon effect. This phenomenon has been widely assumed to function in all well-run propaganda. "Make the voters believe that our man is bound to win and he will indeed carry the election" has been a working principle of every major political party. "But," one may well ask, "what proof is there that the band-wagon effect is at all great, granted that it really exists?"

With the advent of the public opinion polls more than an academic interest has been directed toward the band-wagon effect. For if there is a powerful drive which causes people to agree with the majority of voters, the pre-election polls must disturb the election returns. Congressman Pierce of Oregon, among others, has already voiced his fear that voters tend to climb on band wagons established by straw balloting. And, quite naturally, the promoters of the polls fear that the inevitable result of a widespread acceptance of such a belief will be the legal suppression of their activities. They have therefore been industriously accumulating data to prove the effect non-existent or, at the most, extremely weak. The section below illustrates a few of their counter-attacks.

ILLUSTRATIVE PROBLEMS [13]

Columbus Dispatch data. For more than thirty years the *Columbus* (Ohio) *Dispatch* has conducted reasonably accurate

[13] These data are taken from Gallup, G., and Ray, S. F., *The Pulse of Democracy,* New York, Simon and Schuster, 1940, Ch. 20.

pre-election polls and has built up quite a reputation for accurate predictions. So, when early in 1932 their Ohio poll showed that 65 per cent of the straw ballots favored Roosevelt, one would have expected, if the band-wagon effect is truly important, a continuing swing toward Roosevelt. But, as a matter of fact, nothing of the sort occurred. Instead, when the second poll was taken just before the election, only 51 per cent favored Roosevelt. However, this evidence against the functioning of a band-wagon effect is by no means conclusive. It is at least conceivable that, had the first poll not been taken, Roosevelt would have received only a minority of the votes.

American Institute data. In 1938 the American Institute canvassed Kentucky voters five times in an effort to obtain a preview of what was to occur in the Barkley-Chandler senatorial contest. From the data of Table 55 it can be noted that

TABLE 55 *

THE KENTUCKY BALLOTS

Date	*Barkley*	*Chandler*
	%	%
April 10	67	33
May 15	65	35
July 8	64	36
July 24	61	39
August 5	59	41
	Election Result	
August 6	57	43

* Gallup, *op. cit.*, p. 250.

Barkley's position became increasingly less certain as the election approached. No demonstrable band-wagon effect was apparent.

Shortly after the Kentucky poll came the Georgia Democratic Senatorial Primary. Table 56 shows the distribution of straw votes for the several candidates. Note that if a band-wagon effect had been operating strongly after the September 4th survey, the votes which had been Camp's and, even more, those which previously had gone to Talmadge and McRae (the last mentioned withdrew before the election), should now have

TABLE 56 *

THE GEORGIA BALLOTS

Date	Camp %	George %	McRae %	Talmadge %
Sept. 4	28	52	1	20
Sept. 9	24	52	1	24
Sept. 13	25	46	1	28
		Election Results		
Sept. 14	24	44		32

* Gallup, *op. cit.*, p. 252.

gone to George, the leading contender. Nothing of the sort took place.

In an attempt to scrutinize the issue more intensively the American Institute of Public Opinion, early in 1940, divided the Republicans they had canvassed into two groups—into those who had carefully followed the published data of the public opinion polls and those who had had no opportunity to see these services. Surprisingly enough, the two groups cast almost exactly the same percentages of ballots for the three Republican contenders—Dewey, Vandenberg, and Taft. Apparently it mattered not at all that one group had been previously warned that Dewey was greatly in the lead.

Critical note. The impression must not be given that the lead in the early pre-election balloting is never maintained throughout the race. As a matter of fact, numerous instances could be mentioned where the early leader even increased his majority after a time. Thus, in July, 1937, 67 per cent of the Americans canvassed by the American Institute polling services favored A.F. of L.'s Green to C.I.O.'s Lewis. By October this preference had increased to 78 per cent and by June, 1939, to 80 per cent. But should one hypothesize a band-wagon effect to account for this increase when it was common knowledge that, because of the sit-down strikes, an extremely hostile reaction had occurred at about this time against all things associated with the C.I.O.? However unfair it may seem to some, Lewis received the lion's share of the blame.

Gallup and Rae are probably correct when, in discussing the motive force for balloting, they claim that it grows out of "the impact of events, and the everyday conditions and experiences of the mass of the people" (p. 256). Reasoning such as this does not rule out the possibility that under certain circumstances a band-wagon effect may indeed operate. But if the American Institute data can be accepted as sound it would seem that this alleged motive force in political life is not of sufficient potency to cause alarm.

Level of Aspiration

INTRODUCTION

Level of aspiration defined. In recent years there have been several important studies which have attempted to discover the connections between level of aspiration and subsequent achievement, between what a man hopes to do and what he finally accomplishes. But a basic difficulty has arisen as the investigators, on the whole, have failed to agree on their definitions and measures of level of aspiration. Perhaps the best definition so far offered has come from Gardner [14] who says: "There is, then, one and only one meaning to the term *level of aspiration:* it can only refer to a quantitative indication which an individual makes concerning his future performance in an activity" (p. 66).

To the better understanding of level of aspiration Gardner has contributed a number of rather basic studies, one of the most interesting of which concerns the changes to be expected in a subject's level at those times when he thinks he is improving in skill, losing ground, or just maintaining his score without either gain or loss.[15] Although Gardner conceived his experiment for another purpose, it is this aspect of his findings which will be discussed below.

[14] Gardner, J. W., "The use of the term 'level of aspiration,'" *Psychol. Rev.*, 1940, *47*, 59-68.

[15] See footnote 3.

ILLUSTRATIVE PROBLEM

Gardner experiment described. Gardner's 31 male college subjects were given 20 trials on each of four tasks—card-sorting, digit-symbol substitution, a multiple-choice test of opposites, and cancellation performed while counting backward by 3's. But first they were made acquainted with a percentile scoreboard, a sketch of which appears in Fig. 65. The upper or black marker on the left was originally intended to indicate the score the subject ultimately hoped to reach. However, as that score will not be discussed in this book, no further mention will be made of the black marker.

The instructions were as follows:

> Now . . . I'd like you to set the *red* marker (lower left) to indicate where you expect to stand on your *first trial* in the task. In other words, bearing in mind that the group with whom you are being compared have had 20 trials of practice, what percent of the group do you expect to surpass on *your* first trial? (p. 606)

On each occasion before the subject set the red marker the experimenter set the performance marker to show, as the subject thought, what score had just been made. But, as a matter of fact, this reported level of performance was not the subject's true score, although great care was taken to delude him on this matter. Instead, a prearranged sequence of fictitious scores was presented. This varied a trifle from subject to subject to prevent a comparison of scores which might uncover the deception. The heavy line in Fig. 66 gives the reported level of performance. The broken line records the unusually high aspiration level of one of the subjects. Carefully note the text immediately under the figure. As the unfilled dot records the aspiration for the next trial and is affected by the level of performance which has just been announced (recorded by the filled dot vertically below (the difference between the dots in any vertical line gives the discrepancy score.

Conclusions. Gardner's data indicate that, at least in the experimental situation this set-up has created, the discrepancy

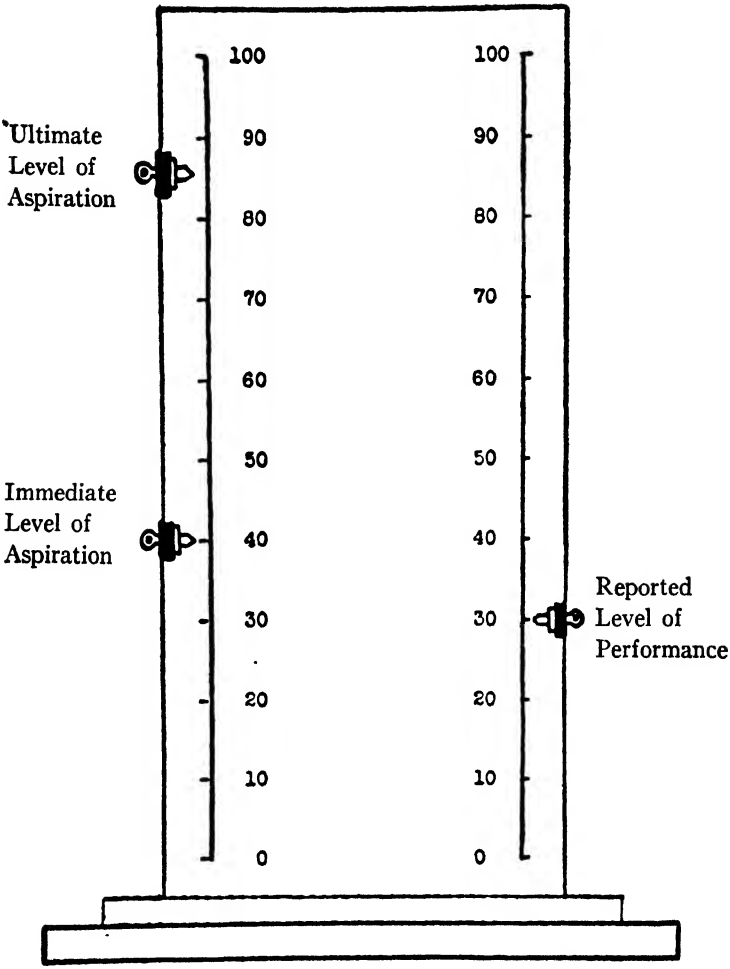


FIG. 65. Percentile scoreboard. Note that (according to these instructions) the first-trial score of the subject is not being compared with the first-trial scores of his fellows but is being compared with the scores which his fellows have attained after twenty trials of practice. This basis of comparison remains the same throughout the test. The red marker is removed.

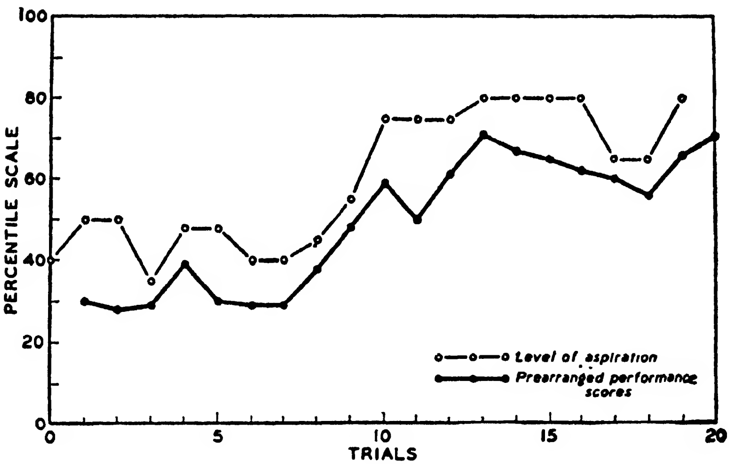


FIG. 66. Each filled dot represents a score reported to the subject. The unfilled dot *on the same vertical ordinate* represents the subject's statement of what he intends to do on the subsequent trial. The discrepancy between the filled and unfilled dots on the same vertical ordinate is the basis for the aspiration level score.

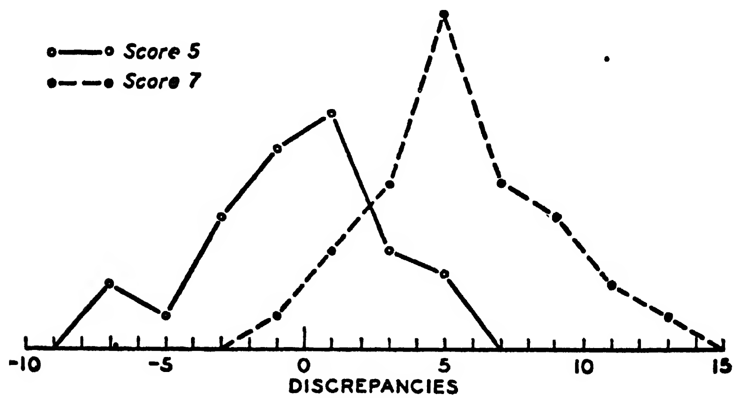


FIG. 67. Distributions of scores 5 and 7. Score 5 is a measure of the discrepancy between aspiration-level and level of past performance when the performance scores are improving rapidly. Score 7 measures the same discrepancy when the performance scores are falling off.

is less when the subject thinks he is improving (trials 8, 9, 10, 12, 13) than when he believes he is losing ground (trials 14, 15, 16, 17, 18). In other words, the subject tends to be slow to anticipate the continuance of a rapid increase in score and is also loath to accept a loss in score. Other studies bear out this generalization. Fig. 67, in which the period of rapid improvement is termed Score 5 and the period of loss Score 7, shows these tendencies very clearly.

Stereotypes

Stereotyping as a timesaving device. No two people behave in precisely the same manner. All Chinese look and act alike only to the Occidental who has made the acquaintance of but a few Orientals. The larger the number of Chinese he closely observes, the more striking become their differences. And so it is with Japanese, Germans, Negroes, capitalists, and maiden ladies. Yet the human life span is so short that even the nonagenarian lacks the time necessary to learn the myriads of discriminations he should make in reacting to his fellows. He is forced to lump people together into ill-defined classes and often to react to the class characteristics instead of to the individual class members. This is most unfortunate as no member fits his class at all perfectly. Moreover the stereotype or class, useful at an early point in its history, often becomes progressively unrealistic until terrible injustices are committed in its name. Thus, the Indians of the sixteenth and seventeenth centuries were cruel savages, or so goes the stereotype. However, even a slight excursion into North American colonial history will disclose the undeniable truth that in the matter of cruelty the conquering Caucasian was at least the Indian's peer.

Stereotypes and reality. The first paper to be reported in this chapter describes a situation whose existence students of social affairs have long suspected: a political party which surrounds itself with a cloak of extreme respectability and thereby gains a rather unreal stereotype can completely annihilate a rival with a "bad" name even though the latter party possesses a reasonably acceptable platform. The moral in this carries over to other fields. The citizen who desires to further a fairly liberal cause weakens his position by joining a "radical"

organization even when the latter is devoted to the fostering of the same worthy project. The person who would aid in improving the moral tone of his city must not identify himself with groups of the "holier than thou" sort. The politician with no program at all to offer can frequently win by mouthing platitudes and waving the banner of the "right" party. To maintain a high social status one must not be a Republican in South Carolina or a Socialist in Vermont.

The method of pooling opinions in determining degree of consistency in social judgments. All of our social opinions reflect in some degree the stereotypes we have accepted. In the second section of this chapter the reactions of typical American college students to the Ten Commandments are presented. Doubtless few will be surprised that the tabu against blasphemy is regarded as of smallest social importance. Yet cannot a time and place be imagined where this tabu would assume greater importance? The order to keep the Sabbath is now regarded rather lightly. But not many years ago violators of the Sabbath were punished far more severely than were those who showed a covetous nature, the tabu against which is now given a far higher rating.

Stereotypes can be found in almost every walk of life. Pooled ratings by both college students and musical experts show Bach and Beethoven to be the most highly regarded of musicians. However, a few years ago Bach had no such eminence, at least among college students. Similarly, our regard for a particular style of architecture changes with time. Our grandfathers loved the "gingerbread" style, while we prefer the "severe." We cannot hope to decide the matter of values solely by direct experience with the art objects. We must depend in part upon stereotypes built up through reading, listening to lectures, and in other social ways.

Occupational prestige. Which of the professions do we regard as of greatest social importance and which of lesser worth? As the third section of the chapter discloses, students in professional schools have rather definite ideas on the subject. Their notions are obviously not inborn but arise rather from

the complex of stereotypes their culture offers them. The jurist is highly regarded in America but not in the Soviet Union. The soldier rates high during periods of international crisis but low during the peaceful times which follow. The small child holds in high regard the fireman, the policeman, and even the garbageman. Now obviously he is not reacting to the specific details of the job of garbage collector, but rather to an unrealistic and childish class description, to the "mental picture" in which the garbageman "gets around" more than do most people, rides on the largest of trucks, receives enormous pay, develops strong muscles, and smokes strong cigars. We all have "mental pictures" of some sort. Occupational prestige, then, is a function of what we have learned—both the true and the false—about the occupations in our culture area.

Definiteness of racial stereotypes. The last section deals with the relation between definiteness of racial stereotype, racial prejudice, and frequency of social contacts between the races in question. It would be a splendid thing if stereotypes disappeared as soon as contacts between the races occurred in large number. But, unfortunately, racial stereotypes persist even after extensive acquaintance. Thus, La Piere [1] has found that the Armenians of Fresno County, California, are commonly considered to be "dishonest, lying and deceitful." Yet the records of the Merchants' Association give them as good credit ratings as those received by the other groups. Moreover, the Armenians apply less often for charity and appear less frequently in legal cases. Prejudices and stereotypes die hard. And, as Katz and Braly [2] show, a racial stereotype need not be very definite to have associated with it a large amount of prejudice.

[1] La Piere, R. T., "Type-Rationalizations of Group Antipathy," *Soc. Forces*, 1936, *15*, 232-237.

[2] Katz, D., and Braly, K. W., "Racial Stereotypes of One Hundred College Students," *J. Abnorm. (soc.) Psychol.*, 1933, *28*, 280-290.

Stereotypes and Reality

INTRODUCTION

Stereotypes as reality viewed through the spectacles of our complexes. Everyday experience teaches us that our fellow men will accede to our wishes more easily if we phrase our requests in such a way as to avoid arousing their "pet peeves," their hostile complexes. No matter how intelligent they may be, they do not behave solely as logical machines. Instead, they view each happening through emotional spectacles. They may condone the criminal actions of Mr. H. "because he comes from such a fine New England family and fine families of course are never anti-social." The fact that the social consequences of a crime perpetrated by one of their peers may be far more serious than one committed by a tramp may not particularly disturb them. They may sympathize far more with a motherless babe than with a child of eight who has just lost his mother even though it is clear that the babe will adjust to the loss more easily than will the older child. Such lack of logic is also common in the political field. A person may be most enthusiastic about a certain social program. Yet, let it get linked to the wrong person or party and his thinking becomes immediately warped.

The Hartmann study. It was to demonstrate this contradiction between actual social philosophy and political stereotypes that Hartmann conducted the study [8] which is described in part in this section. One hundred sixty-eight representative citizens from an agricultural region in Centre County, Pennsylvania, were interviewed at their homes during May and June, 1934. They were offered 20 statements on governmental policy "involving significant social and economic problems" (p. 354). Complete agreement with 10 of the statements which favored such matters as government ownership and management of

[8] Hartmann, G. W., "The Contradiction Between the Feeling-Tone of Political Party Names and Public Response to Their Platforms," *J. Soc. Psychol.*, 1936, *7*, 336-357.

railroads, minerals and land, government insurance, socialism, cooperatives, and the absence of huge fortunes branded the subject as a "radical." Definite approval of the other 10 statements which favored the converse of these sentiments caused the subject to be classified as a "conservative."

The feeling tone attached to 22 real and fictitious party labels was ascertained by presenting the subjects with cards which contained the following directions:

> On each of the accompanying cards you will find the name of a single political party. You probably do not feel the same about each one. *Assuming that the platforms of all these parties are the same, arrange the names on these cards in the order of your liking for them.* Try to answer for yourself the question, "Which name do I like most?" Then ask, "Which name do I care least for?" Finally, place all the remaining party-names in their proper position, according to your *general liking* for them (p. 337).

The anonymity of the responses was guaranteed by the interviewer. The labels *Republican, Democratic, Farmer-Labor* and *Labor* were liked most, while *International, Radical Reform, Technocratic* and *Communist* were liked least.

Figure 68 presents the distributions for 9 of the party labels. These data were obtained from the interviews mentioned above and from the questioning of 173 additional voters from Centre County and 83 adults from Carbon County, a mining district. Note the J-shaped distributions of the well-liked and the disliked labels (see p. 499).

In spite of the clean-cut preference expressed for the more conservative party labels, 55.5 per cent of the responses to the 20 statements indicated support of "a program of socialization with its promise of enhanced status to the working-class population and a corresponding reduction in the power of the privileged groups through a 'redistribution of wealth' " (p. 338). While the women tended to be slightly more conservative than the men, the sex differences were small and not consistent.

Conclusions. Although the majority of the voters were apparently in favor of the specific things for which socialism

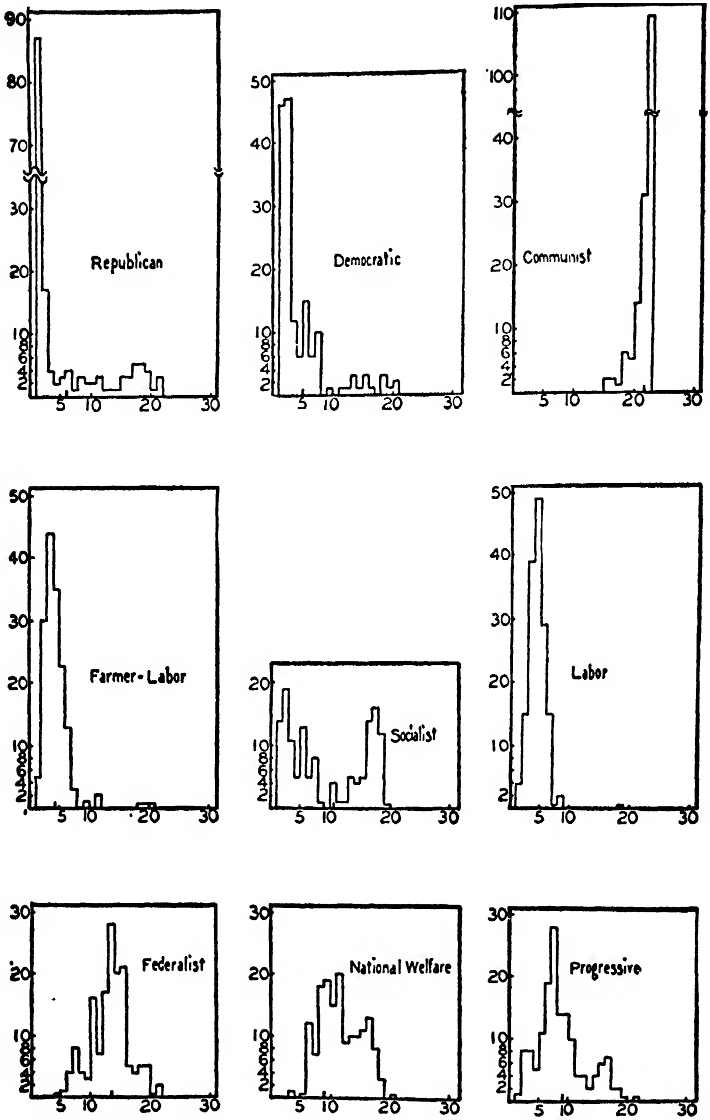


FIG. 68. Sample frequency curves representing the hedonic tone of different party labels according to ratings made by 168 voters.

The small vertical cutting the abscissa indicates the mean for each distribution. Numbers along the abscissa refer to ranks or positions; numbers along the ordinate refer to frequencies.

stands, 61 per cent stated that they did not wish to be oriented toward a socialist society. This phenomenon is in line with that exhibited in the 1940 national elections. The winning New Deal which flourishes under the label of the Democratic party sponsors a program which, only a few years ago, would have been generally regarded as a program of socialization. Yet the traditional upholder of socialization, the Socialist party, polled very few votes. We find ourselves forced to agree with Hartmann that "apparently a sharp discrepancy exists between what most American citizens want and the political channels through which they seek to attain it" (p. 355). They see party platforms through the stereotypes of party labels.

The Method of Pooling Opinions in Determining Degree of Consistency in Social Judgments

INTRODUCTION

Relative standards. Wherever people have fairly similar socio-economic backgrounds we find that they accept common standards. This observation holds true for the arts, for religion, for social practices and for all other phenomena in the field of values. Thus, groups of American college boys agree extremely well when asked to name the greatest men of history in the order of their eminence, or to rank the co-eds of their college in the order of their pulchritude. Of course, Mexican students might be expected to offer slightly different lists and rank orders, while Japanese students would probably behave in a still different manner. That time as well as place is important in the creation of standards is demonstrated by the observation that American youths of a half century ago possessed standards which would clash with those of their grandsons. In other words, methods which disclose the community of opinion existing in a particular field can never yield *absolute* standards. The norms which appear are *relative* to the group studied and will often be quite different at other times and at other places.

ILLUSTRATIVE PROBLEM

Social importance of the Ten Commandments. To demonstrate one type of commonality of attitude which exists among people of similar background, and to present one of the simplest methods for measuring that commonality, data will be presented which bear on the social importance of the Ten Commandments. In the class demonstration to be here described 119 sophomores and juniors from a class in social psychology at the University of Wisconsin were shown the Commandments in the order given in the Bible. This order was to be altered by placing the commandment judged to possess the greatest social importance in first place and that valued as of least importance in tenth position. It might have been anticipated that the term "social importance" should have a variety of meanings for the several subjects, yet the data of Table 57 indicate that there was at least a fair agreement as to its meaning.

TABLE 57

POOLING

Social Importance of Ten Commandments

Topic	*Bible order*	*Ranks* 1	2	3	4	5	6	7	8	9	10	*Sum*	*Rank*
"Monotheism"	I	15			4	12	13	20	22	20	13	795	7
"Idolatry"	II	1	5	3		8	10	15	27	26	24	915	8
"Blasphemy"	III			3	2	1	8	29	25	31	20	952	10
"Sabbath"	IV	2	1	1	2	5	11	20	26	18	33	946	9
"Filial Honor"	V	6	6	10	16	29	27	9	3	9	4	627	5
"Murder"	VI	74	17	6	5	5	4	4	2	1	1	258	1
"Adultery"	VII	3	21	32	27	17	5	3	2	4	5	487	3
"Stealing"	VIII	3	54	28	11	7	5	4	2	1	4	397	2
"False Witness"	IX	11	6	25	29	17	13	7	3	3	5	527	4
"Coveting"	X	4	9	11	23	18	23	8	7	6	10	641	6

N = 119

Consideration of data. Table 57 is to be read as follows. Fifteen subjects regarded the First Commandment as of greatest social importance; four thought it should be rated fourth; twelve, fifth, etc. The sum, 795, was obtained by multiplying each population by its rank and adding across, *i.e.*, (15×1)

$+ (4 \times 4) + (12 \times 5)$, etc. The smaller the sum the more socially important was that particular commandment considered to be. Note that the agreement was far greater for the Sixth (murder tabu) Commandment than for a number of the others, e.g., the Tenth (tabu against coveting).

Size of pool. The size of the pool is one of the variables which influences the magnitude of the agreement. If other variables are held constant, it can be shown that the pooled judgments of two large groups will agree better and more consistently than will those of two much smaller groups. Thus, the rankings by any two subjects picked at random can be expected to correlate at almost any value. That is, they may agree perfectly, just a little, not at all, or they may even completely disagree. But the pooled judgments of any 15 subjects picked at random can be expected to agree fairly well with those of another group of fifteen. This belief is justified by the finding that the pools of two sub-groups of 15, taken from the larger group of 119, correlated at a value (*rho*) of .94. The pooled judgments of two groups, each with a population of 30, yielded a *rho* of .96, while the two halves of the class agreed to the extent of .99. Similar cases where increased agreement is obtained by increasing the size of the populations have been reported by Gordon.[4]

Commonality of background. Another variable which influences the degree of agreement is, of course, the commonality of the background. Two groups of social psychology students from the same college might be expected to agree better than would two such groups from different colleges, or two very different groups from the same university. Thus the pooled judgments of the 119 subjects who served for the above demonstration correlated with the pool of judgments of 52 from a similar class in social psychology at Stanford University at .94; this same large pool, obtained from University of Wisconsin undergraduate students of social psychology, correlated .91 with the

[4] Gordon, K., "Group Judgments in the Field of Lifted Weights," *J. Exp. Psychol.*, 1924, 7, 398-400.

pool obtained from 55 school teachers taking a summer school course in this subject and at the same university. Though these coefficients of correlations are both high, they are below the .99 [5] found between the two halves of the same population. Our demonstration, then, has shown a simple way in which certain of our mores can be examined, and has isolated two of the variables which influence the commonality of opinion to be found among groups of people; namely, the size of the group contributing to a pooled rating, and qualitative differences in the social backgrounds of the groups.

Occupational Prestige

INTRODUCTION

The method of paired comparisons. The problem of occupational prestige is one of values. And values, to the extent that they can be validly quantified, are traditionally measured by one of the major psychophysical methods or some variant. There are a number of these methods, all of which yield roughly the same end results. Yet each is, generally speaking, especially well adapted to some particular need. Thus, to the subject who must judge between a number of stimuli, one of the most satisfactory of the possible procedures is that in which the stimuli are presented to him in pairs. Every stimulus appears once (or sometimes twice) with each of the other stimuli. The rater makes a value judgment by indicating which pair member he considers the better in the quality under consideration—the better liked, the more beautiful, or, as in this instance, as having the greater occupational prestige. The major weakness of this method lies in the fact that the number of possible pairs increases much faster than the number of stim-

[5] While the difference between the *rho* of .99 and that of .91 does not quite meet the usual standard of *statistical* significance (3 times its standard error), it is *probably* significant as shown by the fact that two almost *identical* repetitions of this demonstration yielded very similar results.

uli.[6] Thus, as no judge can be asked to rate more than a limited number of pairs without a possible loss of interest, some other method must be employed if the number of stimuli is at all large.

ILLUSTRATIVE PROBLEM

Coutu experiment. In an experiment on occupational prestige by Coutu [7] 243 (39 per cent) of the engineering students, 142 (30 per cent) of the law, and 202 (58 per cent) of the medical students registered in 1934-35 at the University of Wisconsin were given the following directions.

This is not a test—there are no correct answers. This is a study designed to measure people's attitudes toward certain occupations. Below are listed a series of occupations arranged in pairs. You are to *underline* the one occupation in each pair which you consider the more honorable, the more admirable, or the more worthy of prestige—the one in each pair which you, personally, hold in greater respect or esteem. Judge by your own feelings in each case. If in some instances you are unable to decide, just make a guess, *but do not skip any pair.* Be sure to underline one occupation in each pair, even if you have to guess.

Example: Journalist-Architect

If you respect, admire or esteem a journalist more than an architect, underline *journalist;* but if you respect, admire or esteem an architect more than a journalist, then underline *architect.*

These directions were followed by a list of 190 pairs of occupational names (each of 20 occupations paired with every other occupation).

Data. The votes given to the various occupations by each group of students were counted and transmuted into sigma

[6] The number of pairs can be found by the formula $\frac{N(N-1)}{2}$ where N is the number of stimuli. So, while 5 stimuli can be judged in 10 pairings, 10 stimuli necessitate 45 pairings.

[7] Coutu, W., "The Relative Prestige of Twenty Professions as Judged by Three Groups of Professional Students," *Soc. Forces,* 1936, *14,* 522-529.

scores.[8] A zero score was assumed for the occupation which received the highest number of votes. The other occupations were placed away from this position on a linear scale, the steps of which were in terms of sigma values.

Conclusions. The following conclusions can be safely made:

1. Each of the three student groups gave its own prospective occupation the highest rating.

2. As the typical engineer and lawyer each rated medicine a close second to his own particular occupation, and as the medical students rated no occupation as at all on a par with medicine, first place went to medicine when all three distributions were combined into a single ranking.

3. The engineering list displayed the smallest spread (2.96 sigma) and the medical group by far the largest (5.28). In fact, the position of college professor which came *third* on the medical scale had an identical score with that of the osteopath which ranked *last* on the engineering scale. The range shown by the lawyers was intermediate (3.53) and showed more resemblance to the spread of the engineers than to that of the medical group.

It is clear from the above data that medical students differentiate between the several occupations more severely than do the other groups. The idea of a hierarchy of occupations apparently means more to them than it does to other students. These facts may quite conceivably have an important bearing on the traditional hostility of the medical groups toward socialized medicine; for any move toward such a goal might lower the status of medicine which, at least in the eyes of medical students, appears greatly above that of all other professions.

[8] The sigma score is a relative measure obtained by dividing the deviation of the raw score from the group mean by the standard deviation of the distribution. Sigma scores are held by statisticians to be directly comparable; that is, they express the relative positions a person holds in particular distributions. If, then, he achieves the same sigma score on two tests he occupies the same relative position in these distributions.

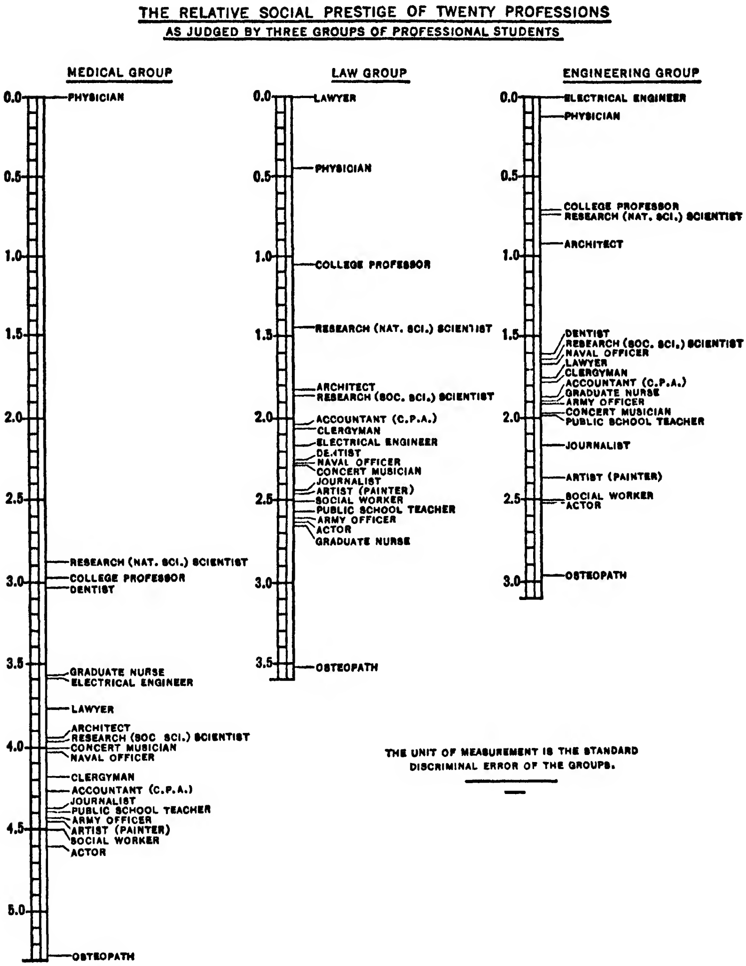


FIG. 69.

DEFINITENESS OF RACIAL STEREOTYPES

INTRODUCTION

Racial prejudice and stereotypes. Only a small minority of us are well acquainted with more than a few members of most races. Many racial and national groups we know solely through the medium of books and other secondary sources.

Moreover these indirect sources of contact are likely to have given us most unrealistic mental pictures. The Turks we brand as a murderous race, for did not they massacre the innocent Armenians? The Dutch must be the greatest per capita soap users as the good housewives do nothing but scrub. And the cultured natives of Borneo are expected to behave in the manner of "The Wild Man from Borneo" of our circuses.

Are our most definite and decided stereotypes associated with those races for which we have the greatest antipathy? Does friendly contact with a racial or national group lessen the aversion ordinarily directed to it? Or may it be that happy personal relations with members of a particular group alter only our own private stereotyped attitudes but not those that we publicly express? (See p. 505.) Katz and Braly have attempted to discover certain of the factors related to the definiteness of the racial stereotype.

Katz and Braly study. In a preliminary study 25 Princeton undergraduates were presented with a list of 10 national and racial groups and the request to "list as many specific characteristics or traits as you think are typical" (p. 284). The list of traits so obtained was supplemented by the investigators until it reached the number of 84. The national and racial (and linguistic) groups to be considered were: Germans, Italians, Negroes, Irish, English, Jews, Americans, Chinese, Japanese, and Turks.

In the major portion of the study 100 undergraduates were asked to select the traits which should be associated with each of the 10 groups. The students were to check the 5 trait adjectives most typical of each race. Table 58 shows the 5 adjectives mentioned most often in each instance.

Even a superficial inspection of the data of Table 58 discloses the fact that Negro, German, and Jewish stereotypes are more definite than the Chinese, Japanese, and Turkish. However, in order to measure the degree of definiteness in a more precise and quantitative manner the following procedure was developed. As there were 100 subjects and each of them checked 5 traits, there were 500 votes to be cast each time a

TABLE 58 *

FIVE MOST COMMONLY MENTIONED TRAITS

Traits	*Per cent*	*Traits*	*Per cent*
Germans		*Jews*	
Scientifically minded	78	Shrewd	79
Industrious	65	Mercenary	49
Stolid	44	Industrious	48
Intelligent	32	Grasping	34
Methodical	31	Intelligent	29
Italians		*Americans*	
Artistic	53	Industrious	48.5
Impulsive	44	Intelligent	47.5
Passionate	37	Materialistic	33.3
Quick-tempered	35	Ambitious	33.3
Musical	32	Progressive	27.3
Negroes		*Chinese*	
Superstitious	84	Superstitious	35.1
Lazy	75	Sly	29.9
Happy-go-lucky	38	Conservative	29.9
Ignorant	38	Tradition-loving	26.8
Musical	26	Loyal to family ties	22.7
Irish		*Japanese*	
Pugnacious	45	Intelligent	47.9
Quick-tempered	39	Industrious	45.7
Witty	38	Progressive	25.5
Honest	32	Shrewd	23.4
Very religious	29	Sly	21.3
English		*Turks*	
Sportsmanlike	53.5	Cruel	54.0
Intelligent	46.5	Very religious	29.9
Conventional	34.3	Treacherous	24.1
Tradition-loving	31.3	Sensual	23.0
Conservative	30.3	Ignorant	17.2

* Adapted from Table 1 in the Katz and Braly article cited in footnote 2, p. 537.

racial group was considered. The measure of definiteness finally set up by the authors was "the least number of traits which have to be included to find 50 per cent of the 500 possible votes" (p. 287). From this it follows that perfect agree-

ment exists whenever 2.5 traits receive 50 per cent of the votes ($250 \div 100 = 2.5$). Table 59 indicates the rank order of definiteness.

TABLE 59 *

DEFINITENESS OF STEREOTYPE

Race or national group	*No. traits required*
Negroes	4.6
Germans	5.0
Jews	5.5
Italians	6.9
English	7.0
Irish	8.5
Americans	8.8
Japanese	10.9
Chinese	12.0
Turks	15.9

* Taken from Table 2 in the Katz and Braly article.

Conclusions. No one can seriously doubt that American college students have had their most numerous social contacts with American nationals. Yet the American stereotype was found to be quite poorly formed. Other studies have shown that of all racial and national groups Negroes and Turks meet the greatest amount of prejudice.[9] But in the table above these two groups stand at opposite ends in the matter of definiteness of stereotype. It is clear, therefore, that definiteness of stereotype is closely related neither to prejudice nor to frequency of social contacts.[10]

[9] Bogardus, E. S., *Immigration and Race Attitudes,* New York, Heath, 1928.

[10] For a comparison with the stereotypes of Negro college students see Bayton, J. A., "The Racial Stereotypes of Negro College Students," *J. Abnorm. (Soc.) Psychol.,* 1941, *36,* 97-102. See also Katz, D., and Braly, K. W., "Racial Prejudice and Racial Stereotypes," *J. Abnorm. (Soc.) Psychol.,* 1935, *30,* 175-193.

Part Eleven

ABNORMAL AND CLINICAL PSYCHOLOGY

By

EDMUND S. CONKLIN
Indiana University

CHAPTER 37

Personal Inventories as Clinical Aids

Introduction. When a human being is stimulated by a question to produce some appropriate and distinctive response, most of the aspects of a typical psychological experiment are present. There is the stimulus in the question; and there is the response, reported more or less correctly by the answer given. Of course there may be much more response to the subject matter of the question than appears in the answer. In addition to the verbal response, there may be also changes of attitude, unobservable changes in muscular tensions, or changes in the patterns of feeling and emotional response. But these additional reactions may be present in any form of human experiment which depends upon the stimulus-response method. When, however, there are several hundred questions and as many answers, there are just as many actual psychological experiments with a single individual used as a subject. This makes systematic questioning an attractive possibility because, if the questions are phrased so as to produce very simple answers in every case, it is possible to record them and to treat them statistically.

If one individual differs significantly from all the others in the experiment because of some brain disease or other organic condition, then the pattern of his peculiarity should show in some significant difference in the general pattern of his answers. If one person in a set of such experimental studies differs from all the others in being a criminal, then his pattern of answers should reveal to a considerable extent the existence of this personality difference. Though it is possible that a single question might be answered by the criminal in a manner significantly different from that of all the other persons involved in the experiment, it appears far more likely that a

peculiar person would reveal his peculiar attitudes if there were many questions used rapidly one right after the other through a fairly long sitting.

The Woodworth personal data questionnaire. Such seems to have been the general line of reasoning which led to the contemporary vogue for the use of long lists of questions in the study of persons suspected of crimes or supposed to be abnormal or different from the socially acceptable in some significant fashion. The interest in this type of work began just before the close of the first world war when R. S. Woodworth of Columbia University, then working for the National Research Council, was given the task of detecting in the large numbers of drafted men those who were likely to be poor soldier prospects because of the probability of (1) their developing psychoneurotic tendencies (mildly abnormal traits such as hysteria), (2) the condition then fairly new and troublesome known as "shell shock" (actually a form of fatigue and psychoneurosis), or (3) any other form of nervous and mental breakdown due to the strain of the new ways of living. Psychiatrists could do a very good job of picking out such persons from any new list of drafted men if they were given the time for long individual interviews in which the conversation would center about the many questions asked by the psychiatrist. But to examine all of the many thousands of draftees in a short time, at least short enough to be of any value to the War Department, was obviously an impossibility. There were not a sufficient number of psychiatrists in the country, nor in the world for that matter. It was quite obviously necessary that some simpler method be substituted and especially that it be one which could be used with many persons at the same time. It was this war need which led Woodworth to utilize for the first time the long list of questions as a kind of psychological test, one which included most of the items commonly inquired about by the psychiatrist. A list of questions was designed with very simple wording; the subjects' answers could be indicated by underlining the words "yes" or "no." Thus, it could be used by many persons at the same time and could be scored by

persons who knew nothing about the subjects or the test or the reasons for the inclusion of any particular question. For the scoring, all that was necessary was some guide by which the person scoring could check the questions dealing with a certain type of problem. These could then be counted and, if the designated answer were that supposed to reveal some abnormality, then the total number of these scores would indicate roughly the degree to which the person answering was in need of special psychiatric attention. In its final form the Woodworth list consisted of 116 questions.[1] This could be distributed to each person in a group, large or small, or to individuals in separate rooms in a hospital, for example.

Obviously the procedure here is notably different from a laboratory experiment, although the list of answers obtained may be treated as though it represented a number of stimulus-response experiments. Some of the questions were so phrased as to make the significant answer "yes," and some so that the significant answer was "no." A list of the significant answers was provided the experimenter who merely had to count on each paper returned the number of such significant answers. With the ordinary class of college students not many items would be answered significantly, about ten on the average; but with persons who were highly disturbed emotionally the number of significant answers might run as high as forty-five or more.

The influence on the answers to such questions of a social situation tending to produce much emotional stress was admirably revealed by Hollingworth in a table which appears in full on page 131 of his book, *The Psychology of Functional Neuroses,* and from which Table 60 has been taken.

If the effect of the armistice upon these responses of persons suffering hysteria and epilepsy can be imagined as affecting

[1] Sold by C. H. Stoelting Co., Chicago, Ill. Also available in Hollingworth, H. L., *Psychology of the Functional Neuroses,* New York, Appleton, 1920, pp. 126; Franz, S. I., *Handbook of Mental Examination Methods,* 2nd ed., New York, Macmillan, 1919; and in Papurt, M. J., "A Study of the Woodworth Psychoneurotic Inventory with a Suggested revision," *J. Ab. & Soc. Psychol.,* 1930, *25,* 335-352.

TABLE 60

EFFECT OF SIGNING OF ARMISTICE UPON RESPONSES OF ABNORMAL SOLDIERS DURING THE WORLD WAR

Question	*Hysterias* During war %	After armistice %	*Epilepsies* During war %	After armistice %
2. Do you usually sleep well? No	55	10	38	15
23. Do you have a great many bad headaches? Yes	72	16	56	32
72. Do you get rattled easily? Yes	36	18	31	30

similarly all or many of the answers, then it can readily be observed that the total score of significant answers would be far lower with the possibility of returning to actual warfare removed than it was during the war. This is one of the facts which such a study clearly establishes.

The Woodworth-Mathews questionnaire. Since the first world war many persons have seen the possibilities of making good use of this Woodworth inventory in studying the personality differences of people in ordinary social life. They have been tried more commonly, of course, upon college students because they are readily available. But they have also been tried upon patients in hospitals for mental cases, upon delinquents and criminals, and upon children of various kinds. It was early apparent that there were a number of questions in the original Woodworth list which were not so appropriate for use with children and that changes in the list were imperatively necessary. The result was what is now known as the Woodworth-Mathews list.[2] This is a somewhat shortened scale with some of the questions edited to make them more appropriate for children. With this form of the list delinquents have been found to produce mean scores significantly larger than those of ordinary children.[3] Others have thought that the schedule really needed revision for use even with ordinary

[2] Mathews, Ellen, "A Study of Emotional Stability in Children," *J. Delinq.*, 1923, *8*, 1-41.

[3] Slawson, John, "Psychoneurotic Responses of Delinquent Boys," *J. Ab. & Soc. Psychol.*, 1925, *20*, 261-281.

adults and so it has been carefully studied and edited for such purposes. One of the best of these revisions can be found in a study by Papurt in which there is not only the original list of questions but also a revised list of 75 questions.[4]

The Laird and House inventories. Persons who are used as subjects for such experiments as these frequently complain that it cramps their style badly to be compelled to answer all the questions with the simple words "yes" or "no." They usually claim that degrees of response should be provided for. Some experimenters have followed this suggestion and, along with certain changes in the questions on the original list, have introduced other and supposedly better ones. Laird [5] produced a series of such inventories for a number of different purposes which have reliability coefficients of around .80 and which provide the respondent a wide variety of possible answers for each question. House [6] produced another which reduced the alternative responses to each question to two "yes" answers of varying degree and one negative answer. The Laird form achieved some popularity for a time but for some reason the House scheme never was so well received.

The Thurstone neurotic inventory. In 1930 L. L. and T. G. Thurstone [7] of the University of Chicago published what has been a widely popular and much used list of 223 questions. In the preparation of this they utilized not only the experiences of those who had preceded them but also adaptations of statistical methods which were somewhat new in this field of work. The Thurstones' purpose in the phrasing of questions for their Personality Schedule was to pick out those students entering the university who might be in special need of the

[4] Papurt, M. J., "A Study of the Woodworth Psychoneurotic Inventory with Suggested Revision," *J. Ab. & Soc. Psychol.*, 1930, *25*, 335-352.

[5] Laird, D. A., "Detecting Abnormal Behavior," *J. Ab. & Soc. Psychol.*, 1923, *20*, 128-141.

[6] House, S. D., "A Mental Hygiene Inventory," *Arch. of Psychol.*, 1927, No. 88.

[7] Thurstone, L. L., and T. G., "A Neurotic Inventory," *J. Soc. Psychol.*, 1930, *1*, 3-30.

advice of the psychiatrist on the student medical staff. Their original purpose was far from being an effort to pick out "neuroticism" or to pick out any special form of abnormality; it was merely to discover persons so maladjusted as to call for early psychiatric attention or at least to make the university advisers aware of their need to be on the watch so as to anticipate and prevent as much trouble as possible. This list of 223 questions is actually a reduction from a much longer list originally tried out, the final items being selected on the basis of their effectiveness in differentiating between persons in the highest and the lowest quartiles of a sample distribution. The statistical reliability is in the middle of the nineties and is thus quite satisfactorily high. A number of the questions utilize words with which even college students are not always familiar. The whole test is thus probably too difficult for the ordinary uneducated citizen to handle with ease, but it was, of course, designed for use with college students.

The score of the ordinary person on the Thurstone schedule will be larger than on the Woodworth because there are so many more possible answers in the whole list. The Thurstones present in their test guide the following distribution of student scores:

TABLE 61

Group	*Scores*	*Description*
A	0-14	Extremely well adjusted
B	15-29	Well adjusted
C	30-59	Average
D	60-79	Emotionally maladjusted
E	80-	Should have psychiatric advice

In using Table 61 as a set of norms it is important to recall that these inventories were filled out by students during their first week in the university, at a time when many might have been still afraid to be completely honest because it might result in failure to be admitted to the university. For this reason the responding students might very likely have revealed less about themselves than they would have under other circumstances. It is known that when college students are asked to

write this same schedule for purely research purposes, with names concealed or nearly so, and without any reference to the problems of entrance to college, the resulting norms are ordinarily somewhat higher than those presented by the Thurstones. This is important because it reveals how easily such a pattern of responses may be altered by temporary social situations.

It is natural to ask to what extent the abler minds among such a group of students may have greater or lower Thurstone scores than the average. Here the correlation with intelligence is very low, approximating zero; indicating, of course, that those having abler minds are just as likely to be maladjusted or neurotic as are the less able. This applies, of course, to persons above the middle of a distribution expressed in terms of IQ, since the lower ranges of an IQ distribution are unable to get into college classes. In the college range of mental ability it is safe to proceed on the assumption that there is no relationship between ability and the healthiness of the social adjustment. The mentally able may be badly adjusted or well adjusted without regard to the degree of their ability.

The Willoughby abbreviated questionnaire. Working with the idea that the number of significant answers to a Thurstone schedule revealed a single something in a personality which might be properly termed neuroticism, Willoughby [8] attempted to present a greatly abbreviated form of the same schedule which would reveal just about as well the relative degree of adjustment or neuroticism. He did this by carefully analyzing responses to the Thurstone list of questions and then selecting those questions which contributed most to the final total score for each respondent. By this means he found and presented a list of 25 questions with the weights to be used in scoring a significant answer to each. With this list it is possible in much less time to obtain results which are comparable with those obtained by the entire Thurstone list. If the Thurstone list reveals a tendency properly termed neuroticism, then

[8] Willoughby, R. R., "Some Properties of the Thurstone Personality Schedule and a Suggested Revision," *J. Soc. Psychol.*, 1932, *3*, 401-424.

the Willoughby list is quite as likely to reveal the same thing to about the same relative degree; but, if the Thurstone list fails to measure any such thing as neuroticism, then the Willoughby list would likewise reveal nothing in this area of value to the investigator.

The Thurstone inventory evaluated. Just what a Thurstone score does represent remains still a much-debated question. All that the Thurstones thought that it meant was the degree of necessity for the student to be referred to a psychiatrist. Others have treated a high score as signifying in the respondent a considerable degree of what has been termed too readily neuroticism. It becomes a difficult question, indeed, if the examiner is asked what the result is when one adds together the significant answers to questions about the personal feelings of a subject, about his organic history and also about his family history. What can the sum of such questions mean? Few thoughtful workers care to answer. Some of the best now insist that a high score on one of these inventories does indicate the probable existence of a degree of abnormality sufficient to make expert attention highly desirable, but that a low score may mean any one of a number of different things.[9] The present writer is not alone in having in his files records of seriously abnormal cases confined in hospitals who have given scores of ten or less, scores that are well within the Thurstone group of those who are said to be "extremely well-adjusted." That they are not well-adjusted but are merely able, for the time being, to answer as they think may be acceptable to the examiner is obvious. Thus the best that can be said at present of a Thurstone score is that if it be high the respondent should be examined by a competent psychiatrist; and, if it be low, then no one can tell what it may or may not mean.[10]

[9] Landis, C., and Katz, S. E., "The Validity of Certain Questions Which Purport to Measure Neurotic Tendencies," *J. App. Psychol.*, 1934, *18*, 343-356.

[10] Persons interested in the significance of very low scores should not fail to read Hathaway, S. R., "The Personality Inventory as an Aid in the Diagnosis of Psychopathic Inferiors," *J. Consult. Psychol.*, 1939, *3*, 112-117.

Examination of the individual questions themselves will soon reveal to almost any student that some of the questions were included with the intention of bringing to light schizophrenic tendencies, that others were designed to reveal tendencies toward those patterns of reaction supposed to be characteristic of the manic-depressive, the epileptic, the psychoneurotic and others with less extreme disorders. But until it was attempted by the present author, no one had systematically presented a scheme for scoring of one of these lists of questions in terms of the different patterns familiar to all students of the abnormal.

It first seemed possible to classify the questions by direct examination, and this was attempted. This arbitrary examination and classification for the purpose of making a diagnostic scoring proved eventually to be a pure waste of time. It became clear that a question which apparently was designed to bring out a familiar schizophrenic trend might not impress the schizophrenic as it did the normal mind. Consequently it gradually became evident that such an arbitrary classification of questions was futile.[11]

The procedure finally adopted for making a diagnostic scoring device involved the observation of what abnormal cases actually did with the questions. For this purpose the responses of large groups of different kinds of abnormal and normal respondents were obtained and carefully tabulated. By this it was possible to find out what questions were answered significantly with far greater frequency by any one group or pattern of abnormality than by others. This soon revealed that there were only a very few questions in the Thurstone list that could be used as diagnostic of tendencies to react as do manic-depressive cases and, because of the small number, that very important pattern had to be left out of the diagnostic scheme.

[11] Page, J., Landis, C., and Katz, S. E., "Schizophrenic Traits in the Functional Psychoses and in Normal Individuals," *Amer. Jour. Psychiat.*, 1934, *13*, 1213-1225; Landis, C., Zubin, J., and Katz, S. E., "Empirical Evaluation of Three Personality Adjustment Inventories," *Jour. Educ Psychol.*, 1935, *26*, 321-330.

But for the schizophrenic pattern there was a sufficient, although small, number of questions which could be isolated and used because the schizophrenic respondents did treat them in a manner quite different from that of any other group. This was also true of a combination of all functionally psychotic (insane) groups in comparison with the normal group used for such comparisons. Hence, another list could be made for indicating similarities to a generally psychotic pattern. Psychoneurotics failed to react to enough questions in such a way that the list could be used as diagnostic for them; but all abnormals taken together showed a pattern of response which reliably distinguishes them from the normal. Thus there was revealed the possibility of a diagnostic scoring for general abnormality, psychotic tendency and schizophrenia, which has proved useful to a number of persons.[12]

Conklin's diagnostic scoring. An example of just how it can be used may be here instructive. Some time ago one of the author's medical friends asked him about an especially peculiar case. The patient was a student who seemed to be obsessed by the idea of need for a surgical operation. He had apparently been able to convince more than one physician of this supposed need for surgical attention. Certainly he bore in the proper places a number of scars which had apparently been originally produced by a surgeon's knife. Aside from this curious desire, he had a number of other strange ways of behaving—enough at least to make the case an unusual one for the more thoughtful physician. The writer suggested to the physician who consulted him that he try the Thurstone list of questions in conjunction with the above-mentioned diagnostic inventory to determine if possible whether the case was more like a psychoneurotic or like a schizophrene. Hitherto the tendency had been to think of the patient as basically psychoneurotic in nature, because of his passion for operations and his other comparable traits. At the physician's request the patient filled out

[12] Conklin, Edmund S., *Three Diagnostic Scorings for the Thurstone Personality Schedule,* Ind. Univ. Pub., Science Ser. No. 6, pp. 25, 1937.

the questionnaire which was returned to the author for scoring. The blank revealed the following pattern of figures:

TABLE 62

SCORES OF A PECULIAR CASE COMPARED WITH NORMS FOR EACH GROUP

	Case score %	*Students* %	*Psychon.** %	*M. D.†* %	*Schizo.‡* %
Abnormal scoring	16	14.9	25.4	29.18	31.00
Psychotic scoring	12	20.7	14.5	31.24	34.36
Schizoid scoring	35	19.65	18.4	20.40	36.75

* The psychon. group were psychoneurotic cases who had hysteria or similar minor mental disorders.

† The M. D. group consisted of persons who definitely had manic depressive psychoses, a type of extreme emotional fluctuation.

‡ The Schizo. group consisted of persons definitely diagnosed as having schizophrenia or dementia praecox, a highly introverted group of psychoses.

It is obvious from this table that the special case was very much like the mean of the schizophrenic group. Apparently he was then developing somatic delusional formations and not psychoneurotic obsessions.

Table 63 presents the best norms that are at present available for use with this form of diagnostic scoring on the Thurstone schedule:

TABLE 63

MEAN POINT NORMS FOR DIAGNOSTIC SCORING OF THURSTONE PERSONALITY SCHEDULE

	Students		*All pathological cases*		*Psychoneurotics*		*Manic-depressive*		*Schizophrenics*	
	Pts.	*P.E.*	*Pts.*	*P.E.*	*Pts.*	*P.E.*	*Pts.*	*P.E.*	*Pts.*	*P.E.*
Abnormal scoring (50 items)	7.45	3.94	14.63	6.04	12.7	5.05	14.39	6.72	15.5	5.85
Psychotic scoring (25 items) .	5.18	2.02	7.31	2.97	3.62	1.44	7.81	2.77	8.59	2.99
Schizoid scoring (20 items) .	3.93	1.61	5.55	2.51	3.68	1.41	4.08	2.33	7.35	2.43

This table presents figures obtained from actual groups of the kinds mentioned. The norms are the results, then, of the actual behavior of the various kinds of abnormal human beings when confronted by the questionnaire and the request that it be filled out. It presents the behavior of normal and abnormal

respondents without the influence of any a priori notions of what individual questions may indicate. As such its validity is probably better than could be obtained by other methods.

Personality types classified. A somewhat similar procedure for the development of a diagnostic list of questions appears as the Humm-Wadsworth temperament scale.[13] It has a comparable diagnostic technique both in its development and in its use. This is, however, based on a different conception of personality. Some years before, a distinguished American psychiatrist, A. J. Rosanoff, published a scheme for the classification of personality types[14] which has long commanded much respect. Humm and Wadsworth utilized this scheme but treated each pattern or type as an elemental factor in every personality. They furthermore assumed that any ordinary personality would be likely to have each of these elements in some measure and a predominance of one of them which would give to that personality its peculiar character. The scheme then was to find questions which would be diagnostic of each of these personality types and then to combine all of these into one list.

The Rosanoff scheme presented the following types: normal, hysteroid, cycloid, schizoid, and epileptic. The last three of these types were easy to obtain for study in institutions for mental cases, but the first two were not so easy to find. The normal is usually doubtful of determination and in this case Humm-Wadsworth found that frequently persons obviously not normal gave so many "no" answers that they would have been in the normal group but for knowledge of their actual behavior. Efforts have been made to bar these out by limiting the range of "no" answers acceptable, but even this is a little unsatisfactory for practical purposes. The selection of hysteroid cases for standardization purposes is even more doubtful. These were possibly more selected in some cases, but as they

[13] Humm, D. G., and Wadsworth, G. W., Jr., "The Humm-Wadsworth Temperament Scale," *Amer. J. Psychiat.*, 1935, *92*, 163-200. See also the description of this scale in Ch. 27, where it was used for the selection of industrial employees.

[14] Rosanoff, A. J., *Manual of Psychiatry*, 7th ed., New York, Wiley, 1938, Ch. 27.

were all taken from among populations of actual convicts, the likelihood of their all being hysteroids is extremely doubtful. Certainly many of them would not represent the hysterical personality as it is met with in the everyday practice of physician and psychiatrist.

The reliabilities of each of these patterns are good. The lowest is .70 and the highest found by the authors is .88. Good as it appears to be, the reason why it has not received more psychological attention is somewhat vague. But failures of recognition do happen. Perhaps the future will reveal this to be one of the best schemes for the determination of personality patterns. On the other hand, the treatment of patterns as elements may prove to be an insurmountable obstacle.

The Bernreuter personality inventory. Another well known list of questions by Bernreuter has rapidly attained a wide popularity. Many studies of various kinds have been made with it. Perhaps its most attractive feature is that it not only does apparently all that the Thurstone schedule will do, but that, with the diagnostic scoring provided by the author, much more can be determined directly from a single response. This diagnostic scoring system presents four different sets of score weightings, one for general abnormality termed neuroticism, one for self-sufficiency, one for introversion and one for dominance. These four traits are naturally quite interesting although, since the first presentation, it has become clear that the correlation between scores for neuroticism and introversion is so close as to indicate that the two sets of weightings are doubtless measures of the same pattern of functions and hence that one of them is superfluous. Reliability correlations have been published for all four of these scorings and all run at .85 or higher, mostly higher. The questions themselves rather closely resemble those in the Thurstone test; although the Bernreuter test is not quite so long it gives a correlation with the Thurstone somewhere in the nineties.

Out of the many studies with this particular list of questions there have come no very certain and conclusive results. There seem to be differences in the findings not yet sufficiently ana-

lyzed. One group of papers, for example,[15] states that the scores on this inventory do differentiate between the normal and the abnormal or, better, between adjusted and maladjusted college students. Such conclusions would indicate its usefulness in many ways and at many times, but it has not been so successful in all studies. In one done under the direction of Landis[16] it failed to distinguish between the normal and abnormal. Here the procedure was much as in the Thurstone study, except that the questions were classified logically instead of empirically; but when a new scoring system was developed upon empirical bases the differentiation was achieved. It thus seems clear that such differentiation is at least possible. It is also almost as clear that responses to the Bernreuter list do not reveal much about other personality traits. The correlation between scores on the Bernreuter and such traits as scholastic aptitude, judgment, verbal discrimination and general information are so low as to be approximations to zero.[17] One is not surprised to learn, for example, that persons with a low or high degree of neuroticism may have either low or high scholastic aptitude, but it is important to know that scores on the Bernreuter do not indicate scholastic aptitude and apparently have little to do with whatever functions are involved in that trait. The same can be said of the other traits mentioned.

The popularity of the Bernreuter seems possibly to have been the motivation for the development of still another list of questions known as the Bell adjustment inventory.[18] This appears to be better adapted for work with high school students than the Bernreuter and Thurstone inventories. For that purpose it appears to be fairly promising, but until it has been

[15] Stogdill, E. L., and Thomas, M. E., "The Bernreuter Personality Inventory as a Measure of Student Adjustment," *J. Soc. Psychol.*, 1938, *9*, 299-315. (Good bibliography.)

[16] Landis, C., Zubin, J., and Katz, S. E., "Empirical Evaluation of Three Personality Adjustment Inventories," *J. Educ. Psychol.*, 1935, *26*, 321-330.

[17] Brotemarkle, R. A., "What the Bernreuter Personality Inventory Does Not Measure," *J. Appl. Psychol.*, 1933, *17*, 559-563.

[18] Bell, H. M., *The Adjustment Inventory*, Stanford Univ. Press, 1934.

used as widely as have the others we shall obviously not know as much about it and what it may do.[19]

The single-word cross-out tests. During all the years in which lists of questions have been used to throw light on the emotional make-up of the individual, there has also been an essentially similar experimental procedure which makes use of reactions to single words. The essential feature in this other method is not that of asking questions which involves, of course, all the variations possible in the understanding of the questions, but that of using a long list of single words. In this method the respondent is told to read these words and to cross out the one in each row of five which is the most unpleasant. The possibility in this procedure of avoiding many of the comprehension errors likely to be present here and there in the responses to lists of questions is obvious, although this procedure is not wholly free from the danger of subjective interpretations of the meanings of single words. But whatever difference there may be, the results as developed by Pressey[20] and others have not proved to be satisfactory, since they have not given sufficiently high reliabilities. Still, the possibilities of the method remain attractive.

The procedure in this cross-out method is simple and should be readily understood. It is to offer the respondent a list of 125 words arranged in twenty-five rows of five each, carefully and designedly arranged. The respondent is told first to go through the list and to cross out every word in each row whose meaning is unpleasant. He is then told to go back and to draw a circle around the word in each row which is most unpleasant. Since there are four such lists, providing altogether for records

[19] A recent effort toward a new list of questions with marked promise for even more successful diagnostic use is reported in the following publications:

Hathaway, S. R., and McKinley, J. C., "A Multiphase Personality Schedule (Minnesota): I. Construction of the Schedule," *J. of Psychol.*, 1940, *10*, 249-254; "II. A Differential Study of Hypochondriasis," *J. of Psychol.*, 1940, *10*, 255-268.

[20] Pressey, S. L., and L. W., "Cross-out Tests with Suggestions as to Group Scale of the Emotions," *J. Appl. Psychol.*, 1919, *3*, 139-150.

of five hundred stimulus-response experiments, it should be of value in revealing the tendency to more or less constant reaction patterns in the particular respondent. They should presumably be reliable, but they apparently are not.[21] Perhaps the possibilities of this method have yet to be realized by careful analysis and experimental check. For the present they must be considered primarily of exploratory value.

[21] McGeoch, J. A., and Whitely, P. L., "The Reliability of the Pressey X-O Tests for Investigating the Emotions," *Ped. Sem.*, 1927, *34*, 255-270; Thompson, L. A., and Remmers, H. H., "Some Observations Concerning the Reliability of the Pressey X-O Test," *J. Appl. Psychol.*, 1928, *12*, 477-494; Broom, M. E., "Note of the Validity and Reliability of the Total Scores Yielded by the Pressey X-O Test of the Emotions," *J. Appl. Psychol.*, 1932, *16*, 681-683.

CHAPTER 38

Experimentally Produced Abnormalities of Behavior

Introduction. Experimentation with animals as subjects has often appealed to the layman as a peculiarly useless type of experimental endeavor. When, for example, he reads that so and so has found that porcupines can learn certain things and that porcupines cannot learn certain other acts, he is prone to remark that it is of little significance because few human beings ever desire to know what a porcupine can or cannot do. He thinks that such experimentation is futile because it appears to contribute little or nothing to the understanding of human behavior problems. But the production of actual abnormalities of behavior in animals which are notably comparable to abnormalities in the behavior of human beings does appeal to the average man as well as the scientist. Here, both think, may be found a promising lead to the understanding of abnormal or mentally diseased human beings. He knows as does the scientist that human mental diseases are but very poorly understood and that for a long time scholars have wanted to study the possibility of producing in animals something comparable with the abnormalities of human beings. When that is possible, we might be within reach of knowledge which may lead to methods of cure far more effective than anything now available.

Most people know that the understanding of many common diseases was achieved by the actual production of such diseases in animals and that experimental work of this sort often led to the development of means of curing or preventing these diseases. Present-day methods of immunization have been produced by this kind of experimentation with animals. Hence

the experiments in recent years on the production of what appears to be abnormal behavior in animals has been hailed with delight.

The nature of human abnormal behavior. Before considering the nature of the procedure and the results of this type of experimentation on animals, one must obviously be somewhat familiar with the nature of abnormal human behavior. Few laymen have ever seen many insane persons or are ever likely to, for obvious reasons. The average citizen, fearing the mental hospital, far too often denies a relative who is mentally ill the best care until it is too late for even the best of methods to achieve what might otherwise have been possible. The fact is that most of those who suffer some kind of mental illness are harmless and need not be feared. Many of them will recover, for a time at least, if removed from the strains and tensions of ordinary living and if provided with good care for their general health. Those whose abnormalities are produced by some organic condition or disease (such as alcoholic or syphilitic invasion of the brain, or the physical disturbances of the menopause) are now treated by fairly well-known and often effective methods.

In every hospital or institution for the mentally sick there are many cases who have no discoverable organic disease. For this reason such cases are described as functional. When they die and a careful postmortem examination is made, nothing can be found that is not also found in postmortems on persons who had never manifested while living any serious abnormality of behavior. Such abnormalities are said to be *functional* instead of *organic*. The latter term is used when there is some known abnormality of the tissues which can be assigned as the cause (syphilis, age deterioration, glandular changes involved in the menopause, industrial poisons, and the like).

The visitor who for the first time sees the patients in a state institution for the insane will observe certain conspicuous kinds of abnormality. If, for example, he should see a large group of patients enjoying the air in front of their residence buildings in the afternoon of a pleasant day, he might be

amazed by the strange antics of certain of the patients. He would very likely see certain of them standing in odd, queer and apparently uncomfortable postures. Others might as likely be seated, but in some queer and unnatural posture held continuously for a long time. Others might be in a fairly normal posture but would be swaying the body, the head, the hands, or be making some regular recurrent movement with the hands or the arms. Upon trying to converse with some of these patients the visitor might be surprised to find a complete refusal to respond to his kindly attempts. Appeal to the attending physician would reveal that these patients were habitually unresponsive, that they refused to respond even to the attendants who were with them constantly, that they even refused to eat and had to be kept alive by tube feeding (a method by which food in fluid form can be poured into the stomach). He would be still further surprised if he should see one of the occasional periods of excitement which these very resistive patients sometimes manifest. In these periods such patients may be very active and often in a very incoherent manner. They appear to be suddenly excited into singing or running or thrashing about in a pattern which appears to be senseless and futile to the sane man. The causes of this behavior have puzzled psychiatrists for a long time. Theories have been proposed and therapies attempted on the basis of these theories, but so far none has proved generally satisfactory and acceptable.

Functional disturbances. The student will have to familiarize himself with still other forms of human abnormality before he can fully appreciate what the experimenters in this field have achieved. There are human beings who exhibit a somewhat wider range of emotional reaction than does the ordinary person. When these people are happy they are happier and gayer and more active than others; but it is equally important to observe that these same persons frequently have periods in which they are blue or depressed and in these states lose their enthusiasms, are disinclined to cooperate and tend to stay in their rooms and mope about. Such are the patterns of behavior in many of those whom we actually know very well,

although ordinarily we are most likely to know them only when they are in their joyous phases. Such people are rather common; but their abnormal counterpart is not to be seen so often, although represented fairly frequently, in the admission files of all institutions for the mentally diseased. Such cases, when genuinely abnormal, are more excited and more exalted emotionally than those we know in normal life, and when they suffer the depressed phase of their pattern they are more depressed and inclined to believe hopeless things about themselves, their friends and the world in general. They think that for them all hope has passed and hence they merely sit, or lie, and cry about it all. In such cases periods of exaltation and of depression ordinarily may be thought of as lasting longer, even up to several months, than do the comparable states in those whom we know have never been suspected of being abnormal but are merely classed as "a bit queer."

It is important to observe in these forms of abnormality the amount and nature of the muscular activity manifested. In the happy states there is much activity, much talking, much eager and apparently useless rushing about; in fact the activity may become so extreme that forcible restraint of some sort may be necessary. In the depressed phase there is a comparable lack of activity. The patient may sit quietly and cry all day long, rubbing the hands, perhaps, as a further expression of the distress he feels. Sometimes this depressed phase becomes so acute that it is properly classed as a form of stupor from which nothing seems sufficient to arouse the patient.

Many functional disturbances of human behavior are not nearly so extreme as those described. They may and often are so slight that they never get to a hospital at all unless, as is often the case, they appear as incidental or superficial aspects of some more profound disturbance. Here reference is made to the little motor disturbances which appear as jerkings or twitchings of some muscle or group of muscles. One sees them as twitchings of the fingers, or the face, of the arm or shoulder, or of some other portion of the body. Sometimes they take the more conspicuous form known as contractures, which may

be mistaken for curvature of the spine, or of inabilities to move the arm or hand in some desired direction. That these manifestations are purely functional can readily be discovered by finding that they disappear completely during the sleep of the patient. There are, it is true, still more extreme forms of muscle disturbance which are catalogued as catalepsies (complete muscular rigidity). These are far less common but nevertheless do appear for short periods in some hysterical persons.

Epileptic behavior. It is also important to observe some of the behavior of persons classed as epileptic. Specialists are still not quite certain just what epilepsy is although, as in other conditions discussed, there are many theories available. Epileptic behavior, however, is here worth considering because of the appearance of allegedly comparable states in animals. The ordinary form of epileptic seizure is sufficiently common so that most persons have probably seen one or more. There is the loss of consciousness and with it the more or less violent contraction of the muscles of the body and the limbs. Careful observers will note that at first these contractions are general and continuous and that in the later phases of the attack they change into a series of sudden, brief or jumpy contractions. As these disappear, consciousness returns, cloudily at first and then with slowly increasing clarity. There is a great variety of these seizures, some very acute, some so mild that they may not be noticed by any except the person who experiences them. Sometimes a violent fit of temper may be substituted for the seizure. In other cases the subject may "run amok" and damage a number of persons in the brief period of excitement. The seizure may be preceded by a period of ecstasy, an experience which such great epileptics as Dostoievski have capitalized in their literary efforts.

For a long time it was more or less carelessly assumed that such abnormalities as those described here were peculiar to man. Only gradually did interest become focused upon actual peculiarities in the behavior of certain animals that might be classed as abnormal or diseased. Some were not at first thought

to be abnormal and the status of others is still uncertain, but discussion of these must be postponed until the facts have been ascertained.

The dancing mouse. An early set of studies in this field concerned the behavior of the so-called dancing or waltzing mouse.[1] From 1898 to 1907 a number of papers and books presented the strange behavior of this little animal and debated the causes of it. The behavior of the dancing mouse is certainly peculiar and interesting. Sometimes it takes the form of placing the feet quite close together and spinning itself around very rapidly. Sometimes it runs around for several minutes in a small circle. At other times the running takes the form of a figure eight. Always these mice are described as being very nervous, very active and as often manifesting quick and jerky head movements.

Explanations of this behavior in the dancing mouse seem not yet entirely clear. Of course some have thought that there might be a defect of the auditory apparatus, but aside from the fact that most adult mice of this species are quite deaf there appears to be no certain explanation possible in these terms. The best guess is that it is a trait which has appeared somewhere in the course of evolution a very long time ago and has been bred by careful although unrecorded selections. The creature seems to have originated in China, where it was treated as an amusing and curious animal, and appears to have been bred for sale because of interest in this dancing trait. Of the hereditary nature of this trait there seems to be no possibility of doubt. The immediate significance, however, lies in the fact that certain mice do manifest such forms of behavior. Are these to be thought of as "insane mice" because they possess a behavior trait so different from the ordinary mouse? Or, are they especially talented? The answer is not readily obtainable.

[1] Yerkes, R. M., *The Dancing Mouse,* New York, Macmillan, 1907, p. 290. (Note especially the bibliography on pp. xix to xxi.) Dice, L. R., "Inheritance of Waltzing and of Epilepsy in Mice of the Genus Peromyscus," *Jour. Mammology,* 1935, *16,* 25-35.

Conditioned behavior of animals. In the justly famous works of Pavlov on conditioned learning there are reports of animals which failed to maintain a normal form of behavior under certain kinds of experimental procedures. Some of this behavior has been thought of as abnormal and comparable to the mental diseases of human beings. The circumstances under which these animal abnormalities occur are important. It must be assumed that the reader is already familiar with the conventional procedure of conditioning animals whereby an originally ineffective stimulus, a bell for example, is made to become effective in producing a given response in the absence of an originally effective stimulus, such as the taste of meat. In these experiments there was often introduced a short time interval between the presentation of the new or conditioned stimulus, e.g., sounding a bell, and the presentation of the old or unconditioned stimulus, e.g., the presentation of the meat. In these experiments it will also be recalled that after the conditioning process had been well established the time interval still held —that upon the presentation of the now effective conditioned stimulus the response did not appear until the conclusion of the time interval. This is known as a delayed response if the conditioned stimulus continues until the response appears, or a trace response if there is an interval between the conditioned stimulus and the response.

In such instances of delay between the conditioned stimulus and the conditioned response it must be quite clear that the stimulus starts the response which for some reason, presumably associated with the learning experience, is held back for a time until the proper interval has elapsed. This holding back, or inhibition, presumably a brain function, was thought to be brought for the first time definitely within the range of experimental study. Of further importance is the fact that in experiments of this kind, when the delay ran up to as long as two or three minutes, the animals often manifested drowsiness and even actually fell asleep. For this reason Pavlov believed that he had been able to demonstrate the actual nature of sleep to be that of a spread of inhibition through the cortex of the

brain. According to his reasoning there are two fundamental and important physiological brain processes, one of excitation and the other of inhibition.[2] In sleep the latter overcame the former for the time being. Presumably, therefore, all states in which consciousness was lost would be instances of dominance by inhibition. Hypnotism was explained in this fashion.

This idea was fascinating because apparently there was the possibility of experimental production of states not ordinarily subjected to experiment. Epileptic seizures might also be explained in similar terms, a violent and intense domination of certain portions of the brain by the spreading process of excitation and with the recovery of consciousness excitation beginning to submit once more to control by inhibition. Speculatively it would also be possible to explain the periods of cloudy consciousness following the seizure in terms of excessive, although not complete, domination by the process termed inhibition. This balance of play between the two processes has been attractive to the thinking of many and may be useful, but for the present problem it is not the most important of the Pavlovian discoveries.

Producing antagonistic response patterns in animals. If by the experimental method of conditioning an animal it was possible to produce a situation in which two antagonistic response patterns were almost equally stimulated at the same time, what would be the result in the animal's behavior? It was possible, for example, to condition some simple response, such as the lifting of the foot, to the appearance of a circle. Obviously such a response would be the result of excitation. If anything but a circle shape appeared, the response was to be refusal to lift the foot. This would as obviously be a response by inhibition. But suppose that the figures shown should be an ellipse and not a circle, would the animal then respond with excitation or inhibition? Suppose, further, that the shape of the ellipse were gradually changed so as to approximate a circle more and more closely, how long would the animal be

[2] Pavlov, I. P., *Conditioned Reflexes*, London, Oxford Press, 1927, p. 430.

able to distinguish between them and at what point would excitation dominate inhibition? This sort of an experiment was tried upon an animal which had been very effectively conditioned to the circle. The result was that when the distinction between the ellipse and the circle became too difficult the animal's former response system broke down; he jumped and jerked and manifested many signs of nervousness so extreme as to be described as "crazy." The animal's usefulness for experimental purposes was ruined, for a time.

What had happened to the animal? Was it true that something comparable to the experience of an insane human being had been produced in the dog? Certainly the behavior resembles fairly closely the extremely excited behavior previously described and found in all textbooks on nervous and mental diseases. Many people have thought that in some of the human abnormalities of behavior the capacity for poise and balance has been ruined by too great tensions. The effort to live in a complicated world has been both too exciting and too inhibiting, and one or the other, or some wild uncontrolled combination of the two kinds of response has been the result. If this view is true and if it is actually possible to produce insane animals by such methods as these, then for the first time in history an experimental approach to the understanding of human abnormalities is possible. Indeed it seems actually to have been achieved. The only necessary further step would be that of trying a comparable experiment upon some human being. But of course this procedure is out of the question, because no normal scientist would dare to produce mental abnormality in a human being even if it were possible to obtain a human being willing to submit to such experimentation.

Maier's rat experiment. The general nature of man's knowledge about experimentally achieved animal abnormalities had advanced little beyond the Pavlov case when a dramatic publication was made by Norman R. F. Maier, psychologist at the University of Michigan. Using similar experimental procedures, he succeeded in producing abnormal behavior in rats. To do this he conditioned rats to jump toward certain stimuli

and then changed the stimulating situation in such a way that their training provided them with no solution; but he forced them to jump from the platform on which they waited by a blast of air or an electric shock. The animal was thus forced to respond without any proper response available. Some of the animals merely jumped vainly against the unyielding stimulus and fell into a net below. But this situation appeared to be so disturbing to the animals that some of them would, after a few attempts, jump or tumble to the floor of the experimental room and then manifest most amazing forms of abnormal behavior.

Maier [3] states that in such cases the rats run wildly about, usually in large circles. Sometimes this excessive activity takes the form of a convulsion or convulsive seizure. Sometimes it appears as a simple jerking of the head or foot, reminding one of the well-known tics in neurotic cases. This first or active phase sometimes gives way to another or passive phase in which the animal appears to be in a state comparable to a coma. Sometimes it fails to respond in any way except perhaps by inhibiting movements. The animal can be picked up, one or another paw may be extended by the experimenter's hand, and when so extended it remains in that position until put back by the experimenter.[4]

This experimental work of Maier's has been widely applauded, presumably in part because of its dramatic nature. Here, far more than in the Pavlovian type of experiment, there seems to be exhibited the actual kind of abnormal behavior which appears in many human beings. Few persons who are familiar with the motor eccentricities observable in any hospital for mental cases can fail to be impressed by the behavior of these disturbed rats. The absurd running about, the convul-

[3] Maier, Norman R. F., *Studies of Abnormal Behavior in the Rat,* New York, Harpers, 1939, pp. 79, 81.

[4] This abnormal rat behavior can be seen by anyone who may have access to the moving picture film presenting these phenomena. The film may be obtained by correspondence with Professor Maier. The book mentioned above contains sixteen different pictures of these rats and presents the behavior abnormalities very well.

sive seizures, the coma, the tics, and the completely unresisting responsiveness to being pushed about (known to all students of the abnormal as waxy flexibility or cerea flexibilitas) are all startlingly like the behavior of mentally abnormal patients. Perhaps the human patient, like the rat, is also pushed to a point where some action is felt to be obligatory but no suitable reaction pattern is available. The animal is made to act when it may be said not to know what to do; just so many human beings appear to break down mentally when they are forced to action at times when they do not know what to do. Maier himself has likened this behavior to epileptic seizures.

Liddell's experiments on sheep and pigs. During the years in which Maier was doing his work on rats other studies had been under way at Cornell University under the direction of H. S. Liddell. These appear to be quite as carefully done and in their results often quite as dramatic. Liddell has worked mostly on sheep but also on pigs in the traditional Pavlovian manner. He has produced much the same kinds of abnormality in these animals. He describes the neurotic sheep as manifesting a state of hyperexcitability, which appears as excessive activity, jumpiness, abnormally quick responses and also as tics, tremors and sudden starts. The capacity for a delayed reaction vanishes. All this looks very much like the reduced inhibition and increased excitability presented by Pavlov, but Liddell is more cautious. He stated as far back as 1929 that the interpretation of sleep as a summation of inhibition could not be demonstrated.[5] His neurotic sheep appear to be permanently abnormal. When dropped from the experimental group the animals appear to improve slowly, becoming gradually more quiet and less nervous. But this improvement is only apparent. If, after months of freedom, they are returned to the experimental laboratory, the old situation causes an immediate return of the neurotic symptoms.

[5] Liddell, H. S., Anderson, O. D., and James, W. T., "An Examination of Pavlov's Theory of Internal Inhibition," *Amer. J. Physiol.*, 1939, *90*, 430; "The Experimental Neurosis and the Problem of Mental Disorder," *Amer. J. Psychiat.*, 1938, *94*, 1035-1043.

A theoretical interest may attach to Liddell's description of the neurotic pig because this animal appears to manifest a dominance of what has been termed inhibition. The pig when neurotic becomes extremely unresponsive. In the study of human abnormalities, failure to respond, to cooperate, to eat, to move when asked, are all classed as forms of "negativism." This is just the way the neurotic pig behaves, and it is not confined to the experimental situation. When allowed to run in the lot with the other animals a neurotic pig runs alone, is ugly toward the other pigs, and may even attack the attendant.

Similarity of abnormal behavior in animals and human beings. Certainly there can be no question of the startling similarity between the abnormal behavior of most of these animals and the abnormal behavior of human beings. Apparently all neuromuscular creatures, man included, function in much the same way and become abnormal under similar circumstances. Behavior resembling abnormal excitement was manifested long ago in the Chinese mouse, and the Chinese, enjoying the curious behavior, proceeded to breed for it with the result that we now have the far-famed dancing or waltzing mouse. In this there seems to be a breed of animals manifesting a curious trait which may be comparable to a strain of human beings in whom some abnormality has become fixed and transmissible. Why the movement of these animals should be in circles is still a problem waiting for an adequate explanation. It is possibly illuminating to recall that when Maier's rats became neurotic they also tended to run in circles. Perhaps this circle running may be merely something produced as a consequence of other and not abnormal features of the rat, though features not found in other forms of animal life. But the solution of this must await more study. Excitable animals and those manifesting excessive inhibition are also comparable to human cases in any mental hospital.

Lush's experiment with goats. Students of epilepsy in humans have long reported seizures in many other forms of animal life, in cats, dogs, oxen, pigs, poultry, rabbits and canaries. More recently a condition thought to be comparable

has been reported in goats.[6] When the goats were surprised or frightened they became completely rigid, could be turned over or moved about as one could move a wooden image of a goat of the same size. The suggestion that this state in the goats is comparable to epilepsy is possibly a little extreme or ill-advised. There is, of course, the possibility of likening it to the period of rigidity in the epileptic, but it is also quite as comparable to those periods of negativism reported frequently in human abnormalities. One wonders, of course, if the negativistic human being is similarly surprised or frightened. Certainly some of the persons suffering such states of negativism manifest pupillary enlargement which would imply some disturbance of the emergency emotion mechanism of the body.

Fascinating as all this experimental work is, and promising for the future as it appears to be, one must nevertheless observe it all for the present with an attitude of suspended judgment. Will it prove eventually, as experimentation goes on, to be a basis for the controlled production in animals of abnormal states reliably comparable to those in human beings? Is it likely that by studying these abnormal animals, methods for the elimination or cure of their abnormal states may be discovered? Can these methods be subsequently transferred to human beings? These are but a few of the questions we should like to have answered immediately, but unfortunately no one can do more than hope and encourage the continuation of such experimental studies.

[6] Lush, J. L., "Nervous Goats," *Jour. Heredity,* 1930, *21,* 242-247.

Experiments with Clairvoyance or Extra-Sensory Perception

Introduction. Recent years have seen a revival of interest in telepathy and clairvoyance (obtaining knowledge at a distance from other minds or objects by other than the usual means of sensory stimulation). Experimental methods are now being used for its study where previously only theory and anecdote had been the source of information. Results from these experiments have been so startling that they have attracted the attention of skeptical scientists as well as the general public. The experimental problem is now to verify what has been reported or to discover wherein the error lies.

The story of telepathy and clairvoyance is a curious one and on the whole very unconvincing. For a very long time the idea had been current that at times and under certain favorable circumstances it was possible for one person to become aware of the thoughts that were in the mind of another. A person may find himself some evening thinking a great deal about a long-absent friend and a few days later receive a letter from that friend indicating that each was thinking of the other at the same time. Two persons walking together may start to say the same thing at the same moment. A very religious person may feel prompted to pray for his friend, a missionary in some remote land, only to find subsequently that at the very moment of the impulse to pray his missionary friend was actually in some very great danger. Such episodes as these have been collected in vast numbers and by some have been accepted as proof of telepathic communication. Others have thought that the power, whatever it may be, could be used also to perceive that which was not within the range of sensory excitation.

Writings in closed boxes have been read by persons endowed with this power, or so it has been claimed. Objects in an adjoining room have been described without direct sensory contact. This form of the power is designated either as telesthesia or, more often, as clairvoyance. The other and allied power of moving objects at a distance with no form of physical contact has been claimed by some and is known as telekinesis, but it has never attracted as much attention.

Anecdotal proof of telepathy. Proof for the existence of any power or ability by the mere collection of alleged instances of its manifestation is said to be based upon the anecdotal method. So far most of the evidence for telepathy has been of the anecdotal variety and is therefore not very convincing. Anecdotes are subject to all the errors of casual observation and distortion by much retelling. It is also very likely that in using the anecdotal method only positive cases are related and preserved; the many negative cases are usually neglected and forgotten. Thus such alleged proof has never been very convincing to the scientifically trained.

A few years ago J. E. Coover, then of Stanford University, subjected many of these claims for the existence of telepathic or clairvoyant means of obtaining information to a most rigorous testing. He found, even with persons who claimed to be "psychics," that they could give results no better than could be calculated according to the mathematics of chance. Obviously, unless such efforts were rewarded with results indicating something more than chance factors present, no one could safely assert the existence of either telepathy or clairvoyance. Because of these findings of Coover's and the fact that, aside from his work, the bulk of the evidence was of the anecdotal variety, psychologists have generally rejected all the claims of those who believe in telepathy and clairvoyance. Although demonstrations in cheap theaters and side-shows often appear to be quite bewildering and for the moment beyond explanation, it has nevertheless been held that such so-called demonstrations must be based upon trickery of some sort. If the persons making the demonstrations have really discovered and

developed to commercial proportions this additional human capacity, the wonder always is that they do not put it to better commercial uses than that of occasional demonstrations in vaudeville theaters. The inevitable consequence has been that scholars believed the whole thing to be based upon fraud, coincidence, accident, and misunderstanding.

Rhine's experiments with telepathy. In recent years the trend of thinking on this subject has been interestingly arrested, for the time being at least, by certain experimental work reported from the laboratories of Duke University done under the direction of J. B. Rhine, and sponsored by no less a person than William McDougall. These experiments have suggested to some that both telepathy and clairvoyance are powers which may be demonstrated and experimented with in the psychological laboratory, even in one's own home for that matter, because very little apparatus is required. The results obtained are presented as indicating something far beyond the influence of mere chance and are, therefore, impressive.

The equipment used is a set of twenty-five cards, of the size and general appearance of the ordinary playing-card except for the face. In place of the playing-card face, the cards are printed with a single symbol on each—a star, a square, a cross, a circle or three wavy lines. These are known as Zener or ESP (for Extra-Sensory Perception) cards.[1] There are five of each of these designs in the pack. Besides these cards, for most of the experimental procedures, only paper and pencil for recordings are necessary.

Simple ESP procedures. Procedure for the experiment varies greatly. It is not as precise and as carefully reported as the procedures ordinarily found in the psychological laboratory (to be discussed later) but it is very easy to acquire, at least in some of its forms. Ordinarily only two persons are involved, the experimenter and the subject or recipient. Every effort is supposed to be made by the experimenter to get the

[1] Ordinarily available in the larger stationery stores or wherever games are sold. If not readily obtainable, these cards and also a book of directions may be procured from the publishers, Farrar and Rinehart.

subject into a contented, cooperative, interested and even happy state of mind. Then, in one form of the procedure, the subject or recipient seats himself in a comfortable chair and relaxes. His eyes may be open and he may be looking in any direction (even at the experimenter). But the subject is told to relax and to report upon signal which of these cards comes into his mind. The experimenter seats himself at an experimental table with the cards in front of him and at his hand a record sheet. The cards are shuffled thoroughly and then cut, sometimes with a knife, for the division. The pack is then placed face down on the table. The first card is picked up held in the hand and looked at steadily by the experimenter. The signal, a tap or click or whatever happens to have been arranged, is given and the subject reports the name of the card which comes to his mind. This is recorded. Then the next card is picked up and again the procedure is repeated. This is done until all the cards have been used once. Then a record is made of the order in which the cards were used for subsequent comparison with the order reported by the subject. Obviously, if there is any such thing as extra-sensory power of communication in this experimental procedure, there is the possibility of either telepathy or clairvoyance or both. The subject may be influenced in his guesses by what the experimenter is thinking about; and likewise the subject might read directly, by his clairvoyant power, the nature of the face of the card attended to by the experimenter.

The DT method. Another form of the experiment is designed to eliminate the possibility of telepathy and to direct the emphasis upon clairvoyance. In this procedure, known to these experimenters as the DT method or Down Through method, the procedure is to place the shuffled and cut pack of cards on the table face down. Then the subject proceeds to call off the cards as he thinks of them without anyone touching the pack until he has called the full twenty-five. The effort is thus to read directly the face of each card as it lies in the pack. When this called list has been completed, the actual arrange-

ment of the cards is recorded for comparison with that of the subject.

Various methods used. Many variations of the above have been developed for one purpose or another.[2] Sometimes five cards, each with one of the five symbols, are laid on the table in front of the subject as guides. Then, after the usual shuffling and cutting, the subject takes the pack, holding it face down, and attempts to distribute the cards each to its proper pile according to the guide cards laid on the table. Of course, he does not look at the faces of the cards but depends upon his clairvoyant or extra-sensory perception of what is on their faces. Again a record is made of the hits and misses. Sometimes the cards are placed each in a separate envelope to prevent any detection of the nature of the card face from other features of the card, such as irregularities in the surface due to the printing. Sometimes a box-like shield is built and placed in front of the experimenter and the cards are held constantly under this shield. Sometimes in the distribution experiment the subject does not touch the cards at all, but merely points with a wand held in the hand. Whatever the procedure variation may be, the critical item is always the degree to which the arrangement of the cards as reported or made by the subject corresponds to, or differs from, the actual arrangement of the cards as produced by the shuffling and cutting.

Some forms of this experiment have been done with a greater distance between the experimenter and the subject. It has been done even with some miles between the experimenter and the subject. Drug effects upon this alleged human ability have been tested by giving the subject caffein or sodium amytal and observing what effect the drug had upon the scores achieved in comparison with scores made by the same subjects without the drug.

[2] The student should see the guidebook for these: Stuart, C. E., and Pratt, J. G., *A Handbook for Testing Extra-Sensory Perception,* New York, Farrar and Rinehart, 1937, pp. 98. See also summary of methods in Rhine, J. B., and associates, *Extra-Sensory Perception After Sixty Years,* New York, Holt, 1940, Ch. 3.

The selection of subjects emphasized. One very important feature of the whole experimental procedure has yet to be mentioned. That is the selection of subjects. In ordinary psychological work there is no effort to select the subjects unless one is studying some form of abnormality in comparison with normal subjects. In the ESP studies the subjects are supposed to be otherwise normal. Large numbers are studied, ordinarily college students, and those subjects who give scores which appear to be significantly high are selected as subjects for further experimental work. They are supposed to have the special clairvoyant power, for the time being at least.

The results obtained by these procedures are of course the important features, and it is these results which have attracted so much attention. The mere assertion of telepathy would not have attracted much attention; that has been made many times before. But the assertion of having discovered a power to read unseen cards, supported by experimental evidence with scores far better than could be obtained by chance, indicates at once either that there is something like clairvoyance present or that there is some other more or less constant influence. It is the report of experimentally obtained support for the claims of clairvoyant influence, with what are said to be statistically reliable data, that has been the real cause of so much attention and discussion.

Scores indicating ESP. In the procedures just described, one should by mere chance achieve a score of about five on the average. With five different symbols to guess from, one should by mere chance get one in five right in the course of a considerable number of trials. But Rhine and his associates report much higher scores. Averages often run up to 8 or 9 and sometimes as high as 12 or 15, though this is very far from being true for all subjects. Most persons who attempt to observe as subjects get about the five score and are dropped as not manifesting clairvoyant power. But here and there a single individual will give much higher scores. It is these who are chosen for repeated experimentation. They may be asked to run through the pack a great many times and as each run

means twenty-five attempts or guesses, the total number quickly becomes a very large figure. A thousand or more is easily achieved. Thus here is one reason for the large numbers which make the reports of these studies so impressive.

It is believed by Rhine and his associates that average scores of higher than five indicate some degree of clairvoyant or extra-sensory power. So it has been stated by Rhine and many others in popular publications as well as in scholarly journals. Scholars in other fields, without psychological laboratory training, have taken up the matter and contended that in these high scores there must be a demonstration of a new or hitherto undiscovered power. Distinguished mathematicians have examined Rhine's mathematical procedures and announced that there is nothing mathematically wrong with them. But psychologists have been very slow to accept the results at their face value or to accept the clairvoyant interpretation given to them by Rhine and his followers. Psychologists have been fooled before; they are prone to be cautious now and to make certain that no other explanation is possible. They also believe that there are numerous errors in the procedure and in the data accepted.

ESP criticized. The gist of the criticisms brought by psychologists against the ready acceptance of these claims and these results may be readily summarized. First of all, it is contended that while the mathematics of chance are doubtless quite correct, it is not proper to use the mathematics of chance as Rhine and his associates have done. The mathematics of chance are based on the assumption that in every instance counted there shall be nothing governing the result other than pure chance. Thus in the case of tossing a penny, there is the pure chance that one out of two possibilities will occur, either heads or tails. In the guessing of an ESP card unknown to the guesser and without any cues whatsoever, the guesser has one chance in five of guessing correctly because there are five of them. Multiply this by five, for the five sets of five in the conventional ESP pack, and the result is five chances out of the twenty-five guesses if every guess is like every other one—a

pure matter of chance. To achieve this constancy of the chance situation it would be necessary to shuffle and cut the pack and then to guess the top or the bottom card or some other specified card in the pack. After that guess it would be necessary to shuffle and cut the pack again and to do so for each guess if each guess is to be comparable to every other one. It is on this basis that the mathematics of chance were developed.

Sometimes the procedure of the ESP researchers approximates to the ideal procedure outlined above. In such cases it is fair to compare the results obtained with what would have been obtained by chance. If the results give an average much different from chance, then it is obvious that some factor other than pure chance is involved. There is probably some more or less constant factor operative. It may be clairvoyance or it may be something else. But this procedure has far from always been the procedure of the ESP experimenters. If, for example, the pack is shuffled and cut and then the first card is guessed and the card turned up for all (including the subject) to see and recorded, the next and succeeding guesses would have a greater chance of being correct than five in twenty-five. If each card in turn were shown to the subject after each guess, the chances for correct "guessing" would rise rapidly. If the person guessing were an accomplished card-player, he would doubtless be able to remember what cards had been turned up and therefore be able to know very nearly what cards remained in the pack. The probability of his guessing correctly would advance rapidly until it became far above the original one-in-five or five-in-twenty-five. Then, if the results obtained by such a method are compared with what would have been obtained by chance, with the mathematics of chance calculated on the assumption that every guess is a pure one-in-five chance, the results are of course going to be much higher than the figures pure chance would give. To conclude that such results indicate some constant factor is correct, but to conclude that it is ESP is obviously fallacious.

Sometimes in these experiments a check-up was made at the conclusion of each five cards. This would also increase the

chance for correct guesses in each succeeding set of five. But the comparison was still made with chance calculations based on the assumption of a pure one-in-five chance each time. Sometimes the subject was told after each guess if it were right or wrong. He could then at least recall some of his hits and infer from them approximately what must follow. Again, it would not be fair to compare the results with pure one-in-five chance calculations and to infer that the more frequent correct guesses indicated some clairvoyant or other peculiar power.

Zenith Radio tests analyzed. It is important to point out the basis for still another adverse criticism of the ESP studies which came to light as a result of the much publicized Zenith Radio broadcasting studies. This is the fact that any series of guesses is commonly influenced by tendencies for guesses to fall into preferred patterns. In these Zenith studies five opportunities were given to the listeners to guess which of two items (black and white, for example) was being indicated. In the studio, as was carefully announced, either black or white was first selected by the spin of a roulette wheel and then ten persons sitting together in an adjoining room all concentrated for the period of the experiment upon whichever item had been chosen by the wheel. The respondents were thus making five successive choices each time between two possible items. Presumably each choice should then be a pure chance matter. But fortunately all of this data was carefully examined by a trained and competent psychologist, L. D. Goodfellow of Northwestern University. He has revealed that many respondents showed [3] what might be termed habitual or recurrent or preferred patterns. Thus again any tendency for the subjects of an ESP card test to guess in the order of some preferred pattern would make it inappropriate to compare the results with figures for chance made on the assumption that every guess is an equally pure matter of chance. Variations from

[3] Goodfellow, L. D., "A Psychological Interpretation of the Results of the Zenith Radio Experiments in Telepathy," *J. Exp. Psychol.*, 1938, *23*, 601-632.

chance might quite as likely be due to the patterning tendency as to the alleged ESP power.

Significance of patterning. In the popular DT or Down Through method (described above) of ESP card guessing this patterning is probably a significant explanatory factor. If in the guesses there were some regular tendency to call in a certain pattern, this would have the effect of producing a score considerably above or considerably below the figures obtainable by chance. Such high and low scores do appear, but it must be recalled that in these experiments there is an accepted practice of discarding subjects who make low scores and of stopping experimentation with a so-called "good" subject if that subject is seen to be producing scores no higher than chance. This would bring out the positive rather than the negative effect. Patterning would obviously not be discoverable in a single run of twenty-five guesses but would certainly be influential in a large number of runs.

Use of high average scores. Another phase of the procedure which makes for better than chance scores is the practice of accepting only high average scores for report and study. Only those subjects who make high scores are used. And even then, when these subjects, after repeated experiments, no longer show high scores but slip back to average, they are dropped or discarded on the ground that their special power has faded or disappeared. Calculations are thus being made upon selected results and thus could not possibly be compared to chance results as a means of demonstrating the presence of a special power. This selection of data has been much stressed by the critics and should be noted here as a second source of error. Some have even contended that proper statistical analysis of the Rhine results is impossible because he has never published all of his data.

Lack of independent records. It is also charged with some force that the lack of independently made records for both the experimenter and the observer may at times considerably increase the scores above chance. Apparently in these studies it is not infrequently the practice for the experimenter to handle

the cards and to record the guesses of the subject. Then, with the subject's record in front of him, he proceeds to enter beside the subject's record the actual arrangement of the cards. Here, obviously, there is the possibility of making mistakes, and of making mistakes in the direction of the desired results. Such mistakes have actually been found.[4] No one in making these charges of error has seriously proposed that there has been intentional falsification, but that even the most careful of trained experimenters may make errors in his own favor is well known and is ordinarily carefully guarded against. Certainly every precaution should be taken to remove the possibility of such errors in this work.[5]

Unnoticed sensory cues. Unnoticed sensory cues have frequently entered in to mislead investigators in these fields of study. A number of instances of supposedly intelligent, even clairvoyant animals have been reported which turned out to be merely animals responding to hitherto unobserved movements of the experimenter, which were picked up by the animal as cues for his responses. Naturally, the possibility of such cues has been looked for in these ESP studies. In some of the cards placed on sale commercially, such cues have been found. Sometimes the ink tends to shrink the card over the area of its application. This has produced an indentation visible on the reverse side and even sufficient for reproduction in photographs.[6] That such sensory cues have not been influential in all of these studies must certainly be true, for all sets of cards do not show them and in much of the experimentation the cards have either been placed in individual envelopes which make them

[4] Kennedy, J. L., "A Methodological Review of Extra-Sensory Perception," *Psychol. Bull.*, 1939, *36*, 59-103.

[5] It has been pointed out that sometimes the persons used to do this type of experimental work have been students working for pay by the hour and that their employment depended upon the success they had in achieving significant scores. In such instances there may have been serious temptations to make the kind of record the employer was seeking.

[6] See photograph of these reverse-side cues in Kennedy, J. L., "The Visual Cues from the Backs of the ESP Cards," *J. Psychol.*, 1938, *6*, 149-153.

invisible or they have been handled by the experimenter entirely behind a screen. But such cues have appeared, and there may be others yet undiscovered. Consequently in all future experimentation of this sort the possibility of sensory cues should be completely guarded against.

It is important to observe in this connection that in the Goodfellow studies of the Zenith Radio returns there was evidence that the respondents were influenced by the wording of the announcements made just before the time allowed for the making of the guesses. The possibility of such verbal cues influencing the guesses has also been experimentally demonstrated where the instructions were deliberately varied for comparison with the results obtained.[7]

Inconstancy of high scores. From the point of view of the Extra-Sensory Perception theory it is a curious fact, but nevertheless a fact, that even the best subjects found do not give consistently high scores. After a run of above average scores, their scoring fades back to mere chance and then they may be dismissed. If this be some special power or capacity of certain human beings, it also appears that it is a power which is not constant. How long it may remain in abeyance, or just what causes it to appear and to disappear, has never been made certain. Fatigue is asserted to have a depressing effect upon it; so also is boredom and loss of interest. In one instance while a young man subject was courting a young lady subject, both appeared to manifest high scores, but after the engagement and subsequent marriage their ESP powers failed.[8]

Experimental setting questioned. The experimental procedure used by the ESP experimenters ordinarily occasions amazement or voluble protest or both in the laboratory trained psychologist. In all experimental psychological work it has long been the accepted practice to keep every item of an experi-

[7] Fernberger, S. W., "Extra-Sensory Perception of Instructions," *J. Exp. Psychol.*, 1938, *22*, 602-607.

[8] Here the reader should, if possible, see Rhine, J. B., and associates, *Extra-Sensory Perception After Sixty Years*, New York, Holt, 1940, Ch. 14.

mental setting uniform. The instructions to the subject are usually written out and either read or memorized and spoken. No conversation is permitted other than what is necessary to make certain that the instructions are clear. Throughout the whole experiment the experimenter says nothing that has not been carefully planned as a part of the stimulating pattern. But in the ESP work all is quite otherwise. Conversation and even touches of gaiety are recommended as necessary to get the subject into a cooperative, willing, interested state of mind. Between runs of the cards animated conversation is indulged in and even encouraged. Sometimes the experimental work is done in odd places, in a parked automobile, for example, out in the country. Such procedures obviously produce patterns of stimuli which cannot be recorded nor exactly reproduced in other settings. Subjects cannot be perfectly compared because it is impossible to reproduce the experimental setting with fidelity.

Attitude of experimenter criticized. It is frequently contended that an important source of error in most of this ESP research is to be found in the attitude of the experimenter. In ordinary scientific experimentation the effort is to get the facts, whatever the facts may be; but, in ESP research, the effort sometimes appears to be, not to get the facts, but to obtain evidence of the existence of extra-sensory perception. The whole attitude expressed in the books of the ESP enthusiasts is that of encouragement to achieve such evidence.[9] In the official handbook by Stuart and Pratt there is a chapter headed "Hints for Succeeding"; and one finds such remarks as "the object of the tests has been to get scores that 'beat chance.' "[10] Such an attitude is calculated more for the discovery of what is desired and for errors in the desirable direction than it is for the discovery of facts whether positive or negative.

[9] See Rhine, J. B., *New Frontiers of the Mind,* New York, Farrar and Rinehart, 1937, pp. 275; Stuart, C. E., and Pratt, J. G., *A Handbook for Testing Extra-Sensory Perception,* New York, Farrar and Rinehart, 1937, pp. 98.

[10] *Op. cit.*, pp. 29 and 40.

Rhine results not duplicated by others. Experimenters in other than the Duke laboratory and by others than Rhine and his followers do not reveal such striking results. It can be said with sympathy and some propriety that most efforts to check the results of the Duke group by persons well trained in methods of scientific procedure have not produced results better than chance. Thus they fail to support the contentions of the Duke group. On the other hand, the Duke group contends that ESP is crowded out or suppressed by the rigid and unsympathetic circumstances of the typical laboratory procedure. Some may think this a valid contention; others may think that it reflects unfavorably upon the contentions of those who begin by believing in the demonstrability of the extra-sensory power.

Experiments limited to ESP cards. Finally, it should be pointed out that the attention-arresting results of the Duke group of experimenters appear to be limited to the ESP cards. Very little has been done with other forms of cards or with any other kind of material for the demonstration of clairvoyant powers. One study was done involving an attempt to reproduce drawings by ESP with what the authors thought were positive results.[11] Others are unable to discover in the reproductions any significant similarities. If this ESP power exists as strongly as contended in some persons, it is strange indeed that such persons should not be able to use it for the reading of the cards held in others' hands while playing bridge; but this power does not seem to have been demonstrated. It is also strange that in the Down Through procedure the subjects do not know when they have finished the pile of cards, but must be told by the experimenter that they have over-run the series. One might properly assume that if a person is able to read what any card is in a pack, one should also be able to tell when he has reached the bottom card; but such is not the case. In the handbook by Stuart and Pratt, recommended by Rhine, the subjects are told that in this DT method it is not necessary to count the

[11] Pratt, J. G., "The Work of Dr. C. Hilton Rice in Extra-Sensory Perception," *J. Parapsychol.*, 1937, *1*, 239-259.

guesses because the experimenter will announce when the subject is to stop.

Some or all of these reasons have led most psychological investigators to dismiss the whole ESP problem as unworthy of their attention. They believe that there is no convincing evidence for assuming the existence of anything like ESP. Others have thought that it might be worth while still to do a little further experimentation in the hope of making the sources of error more clear and convincing, although the actual nature of the errors has already been fairly well demonstrated.

References

Carpenter, C. R., and Phalen, H. R., "An Experiment in Card Guessing," *J. Parapsychol.*, 1937, *1*, 31-43.

Gulliksen, H. O., "Extra-Sensory Perception. What Is It?" *Amer. J. Sociol.*, 1937, *43*, 623-631.

Kellogg, C. E., "New Evidence (?) for Extra-Sensory Perception,' *Sci. Mon.*, 1937, *45*, 331-341.

Kellogg, C. E., "The Problems of Matching and Sampling in the Study of Extra-Sensory Perception," *J. Abn. & Soc. Psychol.*, 1937, *32*, 462-479.

Kellogg, C. E., "The Statistical Techniques of ESP," *J. Gen. Psychol.*, 1938, *19*, 383-390.

Kennedy, J. L., "A Methodological Review of Extra-Sensory Perception," *Psychol. Bull.*, 1939, *36*, 59-103.

Rhine, J. B., *Extra-Sensory Perception*, Boston, Soc. for Psychic Res., 1934, pp. 169.

Rhine, J. B., *New Frontiers of the Mind*, New York, Farrar and Rinehart, 1937, pp. 278.

Rhine, J. B., "The Question of Sensory Cues and the Evidence," *J. Parapsychol.*, 1937, *1*, 276-291.

Rhine, J. B., "The Effect of Distance in ESP Tests," *J. Parapsychol.*, 1937, *1*, 172-184.

Rhine, J. B., Pratt, J. G., Smith, B. M., Stuart, C. E., and Greenwood, J. A., *Extra-Sensory Perception After Sixty Years*, New York, Holt, 1940, pp. 463. Bibliography of 361 titles.

Sharp, V., and Clark, C. C., "Group Tests for Extra-Sensory Perception," *J. Parapsychol.*, 1937, *1*, 123-142.

Stuart, C. E., and Pratt, J. G., *A Handbook for Testing Extra-Sensory Perception*, New York, Farrar and Rinehart, 1937, pp. 98.

Willoughby, R. R., "Prerequisites for a Clairvoyance Hypothesis," *J. Appl. Psychol.*, 1935, *19*, 543-551.

Wolfle, D. L., "A Review of the Work on Extra-Sensory Perception," *Amer. J. Psychiat.*, 1937, *94*, 943-955.

Part Twelve

SYSTEMATIC PSYCHOLOGY

By

ROBERT H. SEASHORE

Northwestern University

CHAPTER 40

Convergent Trends in Psychological Theory

A. THE ROLE OF SYSTEMATIC PSYCHOLOGY

Thus far in this volume we have stressed the progress made by experts who have become highly specialized in their research interests and methods. Experimental psychology always attempts to settle its problems by studying specific instances in which problems occur. But if experimenters stopped after recording their findings in each specific instance they would often have such a mass of details as to confuse both themselves and their readers.

It is nearly always necessary then to summarize the central and deviational trends among these findings in order to see their general significance. The principal methods of summarizing these details employ such statistical devices as the average, the range of individual variations, the coefficient of correlation between sets of details, etc. Such summaries may make it possible for another psychologist to read and understand in fifteen minutes what it took the first investigator months or years to discover.

But suppose one is starting to work on a new problem. How can one find out the most relevant facts, methods, and theories which have already been discovered so as to utilize the knowledge already available and avoid unnecessary duplication of effort? Are there any classifications of psychological data which serve as a sort of filing system with an index for ready reference?

Co-ordination of data. This giant problem of systematizing or co-ordinating the vast amount of data from all the fields of psychology has been approached from several angles. Investigators have always had to organize their findings within the special fields at least well enough to be able to find what

they wanted for the solution of their immediate problems. But in the pioneer exploration and development of their fields the profits from experimentation were so large and the preceding centuries of theorizing without experimental checks had been so unprofitable that most investigators did not take much time off to consolidate their gains into any carefully organized catalogue of their findings.

To a certain extent the founding of psychological laboratories, associations, journals, and collaborative handbooks of information have assisted in the integration and diffusion of knowledge by bringing together in convenient places the facilities for investigation, discussion, and publication of research findings. An important example is the *Psychological Index* which until recently was published annually, classifying the titles of all new articles and books, according to special topics. As more and more publications began to appear, and psychologists could not read all of the first-hand reports, the annual *Index* was displaced by a monthly journal, *Psychological Abstracts,* which gave not only the reference but also a short summary of each report. At present it is proposed to follow the lead of other sciences in having a still further summarization of the constantly increasing flood of new findings in an annual review of only the most important contributions in each field.

Evidently, then, there is a need for some system of organizing psychological data according to major topics, so that psychologists can readily discover what facts are already known, what theories have been proposed for further study and what methods are likely to be most applicable to each type of problem.

The function of systematic psychology. Systematic psychology attempts to meet this need, to classify the major topics of psychology in convenient form so that specialists in all of its various subdivisions may use this filing system for recording their own findings, discovering their relations to other findings, and getting suggestions in the way of theories and

methods for filling out their knowledge of the blank spaces in each area.

A complete system would be an encyclopedia of all the knowledge in psychology. In contrast, a psychological viewpoint or school of thought is only a fragmentary system and usually covers only a small portion of all the major areas which have usually been considered to belong to psychology.

Most of the modern viewpoints in psychology have each attempted to develop a system which is based upon the major interest, methods, and concepts of their own approach. These attempts have had the advantage of organizing each of their own limited fields of interest within themselves, but usually they have struck difficulties in attempting to cover the whole area of psychology.

This type of attempted system-building has led to the anomaly of having six or seven theoretical developments, each claiming to be *the* necessary or most suitable basis for a science of psychology, and denying the importance of most other such claims. Such controversies have received widespread publicity and have done little to build respect for systematic psychology.

Most of these attempts at system-building have found it possible to fit some of the data from other viewpoints into their organization, but other data have not fitted very well, and some data have fitted so poorly that the various schools have attempted to throw it out as not belonging to science of psychology.

In the face of such controversial theoretical developments, many psychologists have chosen to follow no single viewpoint or school but instead to accept an eclectic view which attempts to combine the best features of all viewpoints. Such eclecticism has often resulted in the retention of numerous apparent inconsistencies among the contributions accepted from the various schools. Eclectics have usually preferred merely to suspend judgment on such points until they have been cleared up in the course of further experimental investigations.

In the meanwhile the development of better contacts between workers in various fields of psychology has cleared up

many of the minor inconsistencies and certain general fallacies have been uncovered which seem to explain a great many of the more serious remaining inconsistencies.

In attempting to evaluate the contributions and hindrances of the principal modern viewpoints to a complete system of psychology we have set up a chart which lists in a logical order all of the major problems which historically have received the attention of psychologists. Opposite each such problem or question we have listed the principal contributions and subsequently discovered mistakes of each major viewpoint. The result is a chart which at first is conspicuous for the very large proportion of blank spaces in the columns representing the work of the various schools. Evidently then no single school has supplied answers to all of the major problems, and usually there are only a few of the many schools which have attempted to deal with any one single problem. Closer examinations will show that even where two or more schools have been interested in the same general problem, they usually have not been concerned with the same aspects of the problem and thus could hardly be in conflict since they were not even talking about the same thing.

Instead of considering each viewpoint to be in conflict with the others, it now appears that their differences were largely in the realm of specialization of their interests. Each viewpoint has usually concentrated its interests upon the investigation of only a few items from the list of major problems, and ignored or made little contributions to the rest. It might even be said that no one viewpoint has ever completely covered a single one of the major problems, but rather that it has dealt with a few related aspects of several problems.

Our first point will therefore be that each of the viewpoints represents a rather arbitrary definition of the proper scope of psychology, and that any definition of psychology which included all the major historical contributions to the science would have to include all of the various viewpoints as supplementary aspects of psychology. In spite of the apparent conflict between supposedly divergent viewpoints, each has been con-

tributing significant data to psychology on various problems and has failed principally in not adequately attempting to see how each of them fitted into the picture as a whole.

The scope of psychology. Definitions of the scope of psychology are constantly changing from one decade to the next, each viewpoint tending to define the field as a whole in terms of its own limited current interest. Historically psychology has been just "what psychologists study"; and as various problems approach solution, new investigators move on to newer interests. It is now generally recognized that there are no sharp natural boundaries between sciences and that our present dividing lines are simply matters of convenience to aid in the division of labor among scientists. It follows, then, that any further attempts to define psychology in terms of any special set of interests are scarcely worthy of serious attention. The various special definitions of this sort need only be examined in order to see how they fit together. The first fallacy of the various attempts has thus been that of disputing about *arbitrary definitions of the "proper scope" of psychology.* The second fallacy is one which is closely related, namely, the *failure to recognize supplementary aspects of a problem.* Their supposedly conflicting statements often showed no genuine logical opposition but instead constituted supplementary contributions to different aspects of a problem.

When different schools have actually been talking about the same aspect of a given problem they have often failed to recognize a third fallacy, namely, the mere *differences in technical terminology* for the same general findings. These too have often been mistaken for genuine logical opposition between viewpoints.

A fourth and perhaps most important fallacy is the "all-or-none" acceptance or rejection of a given viewpoint. On the grounds of probability alone, the chances of any viewpoint being practically all correct or practically all wrong are very unlikely. Nearly always it is more reasonable to expect that any viewpoint contains some valid and some invalid ideas. Even if it is only one-half per cent valid, it will eventually be needed

in a complete system, and if only one-half per cent is invalid, that part will eventually have to be weeded out. Furthermore it is often found that statements which may not be literally correct are at least in the right direction and may be revised so as to be acceptable in a scientific system.

According to our theory it should thus be possible to list a relatively small number of major problems in psychology and to show that each of the modern viewpoints has contributed numerous facts and methods which are gradually being fitted together to fill out our systematic outline of the entire science.

Main problems and contributions. Obviously, in a single chapter we can mention only the highlights of both problems and contributions, but these should be sufficient to indicate the adequacy of the plan. We shall therefore begin by listing each problem; then we shall give a brief statement of the contributions of five modern viewpoints, together with an eclectic summary in each case. The divergent aspects of the various schools are amply developed in practically all historical and contemporary texts such as Woodworth's *Contemporary Schools of Psychology,*[1] Heidbreder's *Seven Psychologies,*[2] and Murphy's *An Historical Introduction to Modern Psychology.*[3] We shall therefore stress the neglected aspects of *convergent trends in psychology,* a vast but peaceful series of developments which has so far attracted too little attention.

The present chapter will limit itself to the consideration of five major theoretical problems, the answers to which constitute the general working assumptions of each of the special viewpoints and those of the eclectics.

[1] Woodworth, R. S., *Contemporary Schools of Psychology,* New York, Ronald Press, 1931, 232 pp.

[2] Heidbreder, E., *Seven Psychologies,* New York, Century Co., 1933, 450 pp.

[3] Murphy, J., *An Historical Introduction to Modern Psychology,* New York, Harcourt, Brace, 1932, 471 pp.

B. SOME MAJOR PROBLEMS AND THE CONTRIBUTIONS OF THE VARIOUS "SCHOOLS" OR VIEWPOINTS

Any of the modern viewpoints can be summarized in outline form by the positions which they have taken on these five major theoretical questions. Although these positions have often been stated as very positive assertions, they now seem to have been only working hypotheses as to the nature of the data studied, the particular aspects of each problem which seem to be most important for psychology, and the relations of psychology to other sciences and technologies.

Question 1. *What is the "proper subject matter" of the science of psychology?*

The first viewpoint which we shall consider is that of *existentialism,* or *structuralism,* a school of thought which grew out of English associationistic psychology of the seventeenth to the nineteenth centuries, matured as the experimental psychology of Wilhelm Wundt's laboratory at the University of Leipzig from about 1879 to 1920, and reached its highest development in the hands of Wundt's student, E. B. Titchener, who directed the research of many students at Cornell University from 1892 to 1925. Wundt had participated in a long controversy with Brentano of the University of Würzburg as to whether psychology should merely study mental processes as "contents" of the mind which were found to exist, or whether it should also describe how these processes functioned. In convincing other psychologists of their views, Wundt and his students, who were pioneering in laboratory experimentation, had the advantage over Brentano and his students, who relied more upon everyday observations.

Wundt and Titchener believed that psychology should have a definite *content* or subject matter comparable to that of physics. Wundt himself was interested in, and wrote widely on, many different aspects of psychology, including folk or social psychology and animal psychology; but Titchener[4] later in-

[4] Titchener, E. B., *A Textbook of Psychology,* New York, Macmillan, 1909, 311 pp.

sisted that if psychology were to be considered a science comparable to physics, it should restrict its status to that of a pure science which studies the *generalized content of the normal adult human mind.* Each and every major word in this conception eliminated another large area of interest which had been studied by other psychologists. The word *generalized* eliminated considerations of individual differences except as sources of error. The word *content* excluded the study of function. The word *normal* excluded the field of abnormal psychology. The term *adult* excluded all earlier phases of child psychology, while the term *human* likewise excluded animal psychology. Finally, the term *mind* excluded the study of overt behavior. Titchener still had enough of the central area of psychology left to occupy him and his students for a lifetime, and his studies resulted in a monumental four-volume manual of experimental psychology, together with numerous other volumes, such as introductory texts and volumes on the feelings and thought processes. It was perhaps a wise idea for him to concentrate the work of his laboratory on certain of the older major problems before going on to other fields, but the arbitrary manner in which other psychological problems were ruled out antagonized many other psychologists who wished to study other fields also. Titchener believed that a science should study only things *as they are found to exist,* hence the name existentialism. This excluded all reference to the *meanings* of the processes, which he considered to be a speculative problem in the philosophy of values. He further warned against the common error of confusing descriptions of mental processes with descriptions of the stimuli which set them off (the stimulus error). In keeping with this, the pure science of psychology was to be clearly distinguished from technology or applied science, however valuable that might be in other ways.

A second viewpoint advanced by the *American functionalists* drew not so much from Brentano's idea of function as from the biological evolutionary concept of the utility of functions as stressed by Charles Darwin. The American functionalists, William James, who began teaching at Harvard about 1870,

and a group of men led by John Dewey, James R. Angell, Charles Judd and Harvey Carr who began work at the University of Chicago about 1890, had found studies of the biological usefulness of vision, etc., to be very interesting and worthy of scientific study. Largely because the legitimacy of this research interest had been challenged by Titchener, the Chicago school proceeded to develop their position in detail, though never to the exclusion of other viewpoints. Members of the Chicago school were very broad in their definition of psychology, Carr's definition being: [5]

> Psychology is primarily concerned with the study of mental activity. This term is the generic name for such activities as perception, memory, imagination, reasoning, feeling, judgment and will. . . . Mental activity is concerned with the acquisition, fixation, retention, organization, and evaluation of experiences and their subsequent utilization in the guidance of conduct. . . . The type of conduct that reflects mental activity may be termed adaptive or adjustive behavior.

All of this time psychologists had been plagued by the suspicion that the subject matter of psychology, mental processes, was subjective, that is, open to observation by only the one person whose experience was being studied, and therefore could not be verified independently by other observers, as could the subject matter of older physical and biological sciences. This supposed lack of independent verification seemed to be a great stumbling block to progress, and psychologists constantly sought more clear-cut methods of study to replace introspection.

As we shall see later, the problem was not as serious as it seemed, but it did lead to a third important movement, called *Behaviorism*, or the study of overt behavior, which first became prominent with the work of J. B. Watson, about 1911, who had recently left the University of Chicago to join the faculty of Johns Hopkins University. Watson, who had experimented in animal psychology, was impressed with the fact that in this

[5] Carr, H. A., *Psychology:* A Study of Mental Activity, New York, Longmans, 1925, 432 pp.

field independent verification of observations was possible, just as in the other natural sciences where as many observers as desired could watch the behavior of the animals and compare notes. This fact, together with his disgust over a long-drawn-out controversy between Titchener's and Kulpe's groups of investigators on introspective evidence as to the nature of the higher thought processes, led Watson [6] to renounce introspection entirely as a method and to propose that psychology be defined as the study of *behavior*. All mentalistic terms were to be cast aside, with the hope that this temporary inconvenience would be more than justified by the eventual building of a much more adequate system of psychology which would give it a clear title to scientific standing along with the biological and physical sciences.

Behavior had been studied by earlier psychologists, and even stressed to some extent by Pillsbury, McDougall, Parmelee, and Meyer, about 1900. Watson, however, was the first to propose behavior as the exclusive subject matter of psychology; and, as might be expected, the proponents of the older viewpoints attacked the narrowness of Watson's definition immediately.

A fourth viewpoint, the *psychoanalysis* of the Austrian psychiatrist, Sigmund Freud, had begun in the first decade of this century, but had not developed to the extent of attempting to cover the whole field of psychology. It consisted primarily of a body of information and of methods for the analysis and treatment of mental disorders. From the standpoint of definitions of psychology, it is interesting mainly because it emphasized the importance of studying *unconscious* mental processes, a point which had cropped up often in the history of psychology, but which, because of its vagueness, had usually been avoided as impossible or unworthy of scientific study. Nevertheless, difficult or not, if Freud is even approximately correct in his suggestion that the unconscious mental processes are analogous to the 90 per cent of an iceberg which is below the

[6] Watson, J. B., *Psychology From the Standpoint of a Behaviorist*, New York, Lippincott, 1919, 429 pp.

surface as compared to the clearly perceptible 10 per cent above, their study can no longer be ignored in general psychology. Psychoanalysis is primarily concerned with the nature and origin of the motivation behind our thinking and behavior, particularly as it leads to the development of mental disorders. It is a clinical approach, not particularly concerned with its systematic relation to psychology in general.

A fifth viewpoint, *the Gestalt movement* led by Wertheimer, Koffka and Köhler in Germany, began about the same time as Behaviorism, but was little known in America until scientific communications were reopened several years after the first world war. The German term "Gestalt" is usually translated "configuration," and refers to the configurations or patterns in which, they hold, experience or behavior occurs. Gestalt psychology is also called organismic psychology, by analogy with organismic biology, both stressing the action of the *organism as a whole.* An outgrowth of this earlier Gestalt psychology is Lewin's topological or field psychology which attempts to analyze patterns of experience and behavior into vectors rather than into elements.

These vigorously presented viewpoints may be said to represent scouting parties, pushing out into new fields ahead of the main body of psychologists, who are usually eclectics attempting to sift out the valuable contributions from each new viewpoint or investigation and attempting to orient these in some broader outline of psychology. If one simply classifies new contributions from various fields, the arrangement may at first appear to be a mere series of piles of facts, theories and methods. Some of the piles can be seen to be logically related, while others show no apparent relation to one another, and still others seem to be logically contradictory.

As compared with the few bold outlines of any of the special viewpoints, the early stages of any eclectic system appear to be almost as disorganized as a jigsaw puzzle. As details are added and fitted together in various areas of investigation, the outlines of the eclectic system begin to clear up. Eventually we may get some rough sort of a system from the mere accumulation of

facts from separate investigations and their observed interrelations. This, however, would be an extremely slow process, and needs to be supplemented by active theoretical investigations which try to classify psychological findings in various ways, searching for the system which will adequately account for the most facts with the fewest and simplest assumptions. An example of this type of system-building is the one mentioned previously, in which the principal assumptions are as follows:

1. The classification of the subject matter of various sciences is largely a matter of convenience to aid in the specialized division of labor among scientists.
2. Most of the areas which have been investigated by well-trained psychologists may be considered to belong at least in part to the subject matter of psychology. This recognizes the frequent occurrence of joint interests with other sciences.
3. A great many of the supposed discrepancies between the contributions of various viewpoints are to be accounted for in terms of the four major sources of confusion in our thinking, as previously mentioned: *arbitrary definitions* of the field, mere *differences in emphasis on various aspects of a problem,* mere *differences in technical terminology* which really are equivalent to one another, and the tendency to evaluate viewpoints or separate contributions on an *all or none* basis rather than by more careful sifting of strong and weak points, together with their possible further clarification.

In keeping with this approach, an eclectic definition of modern psychology might be stated as follows:

Historically, psychology has come to include the study of sensory, affective, motor, and intellectual processes in all of their aspects, e.g., general vs. differential, qualitative and quantitative, pure vs. applied, normal vs. abnormal, elemental vs. organismic, content vs. function, etc. Specific schools have simply emphasized a few aspects of some or all of the four processes but no single school

has covered all major aspects of all four processes. The viewpoints *supplement* one another with relatively little overlapping.

Question 2. *How are mental processes related to physical processes?*

This major philosophical problem is also important for psychology because it considers whether mental processes are independent of physical processes, merely parallel to them, caused by them, or merely one aspect of a unitary psychophysical organism. Beginning with the experimental psychology of Wundt, most psychologists wisely decided that the best thing to do about this problem was to adopt some one answer as a working hypothesis and then keep testing it as new facts arose.[7]

Thus the existentialists adopted the hypothesis of psychophysical parallelism, which assumes that mental processes run parallel to physical processes (brain action, etc.), but are fundamentally different in nature. They did not explain how they could run parallel without any causal relationship, but simply recognized this as a point to be settled later. The important thing was that they quit debating in the philosophical manner which had made so little progress over a period of two thousand years, and started acquiring experimental facts which eventually have done much to clear up this problem.

The functionalists held no single hypothesis exclusively, but were inclined to adopt the double-aspect explanation, that mental and physical processes were simply two different aspects of the functioning of the same organism. Their work helped to show that our so-called "mental" processes (seeing, feeling, thinking, etc.) actually involve the functioning of the whole body, including the muscles, and not merely the sense organs and brain.

The behaviorists, wearied by the debate over this problem, simply ignored it by stating that all we can study scientifically is the behavior of organisms. Terms such as visual response

[7] Ogden, C. K., in his *Meaning of Psychology*, New York, Harper, 1926, has outlined the nature of each of these hypotheses quite clearly.

were substituted for "seeing," but little attempt was made to translate all such mental terms into physical terms until very recently. Behaviorism said that there is only one kind of ultimate reality, and that is physical, involving the electrons, protons, etc., studied by the physicists. Biology and psychology simply involve the study of more complex forms of the same physico-chemical nature.

The psychoanalysts were not particularly interested in the physical aspects of mental life, and proceeded to deal with both conscious and unconscious mental processes with little regard for their physical relations.

The Gestaltists have proposed a principle of isomorphism; i.e., there is a parallel in the pattern of brain action for every pattern of experience. This is perhaps nearest to the double-aspect theory held by the functionalists, but has never been specifically classified by the Gestaltists themselves. They have attempted to work out rather elaborate theories of the way in which the nervous system functions during vision, learning, etc., but there are still too few facts known about functional neurology to make feasible any definite evaluation of their theories. We shall merely point out that the Gestaltist and organismic theories assume patterns of electro-chemical fields in the brain which are transposable to different parts of the brain. In contrast to this, the older theories supposed that there is a high degree of fixed specialization among nerve fibers, so that they act like partly insulated wires in a central telephone exchange.

The eclectic answer to this question has probably been best stated in the double-language theory of C. K. Ogden,[8] which holds that mental and physical descriptions are all part of a single system, but that unfortunately, in the early days of the isolated development of science and philosophy, two different sets of terminology got started, and that we are just now beginning to co-ordinate these terms. After reading Ogden, one can see that the earlier mind-body theories were often partly

[8] *Op. cit.*

right, but fell down on specific points which are gradually being straightened out as workers from the various sciences cooperate on problems of mutual interest.

Question 3. *Do psychological processes function in essentially the same way as inorganic physical processes (the mechanistic view) or is there some other characteristic of living things, which is entirely different from the inorganic physical processes (the vitalistic view)?*

Existentialists, functionalists, and behaviorists clearly favor the mechanistic view, believing that there is a continuous development from the laws of physics and chemistry on up through biology to psychology and the social sciences. The term "mechanistic" means "machine-like," and with the progress of physiological psychology, many similarities have been found between psychological functions and those of physical and chemical apparatus.

The Gestaltists have objected to what they term mechanistic views, but the particular point to which they seem to object most is the assumption of an analogy between the behavior of organisms and the operation of *simple, rigid* machines, such as levers, pumps, etc., which can be operated in only one fixed way. But the modern mechanists would point to analogies with such complex apparatus as photoelectric cells, radio receiving sets, and apparatus used in organic chemistry; in other words, to a whole series of very complex machines. Whether Gestaltists will agree with this newer interpretation of mechanism is not yet known.

Psychoanalysis states that all mental processes, conscious and unconscious, are determined by fixed principles, not yet completely known. Even the phenomena of dreams and the symptoms of mental disorders have an origin which can often be discovered by the use of psychoanalytic principles or laws. This viewpoint is called determinism; it is essentially the same thing as mechanism, but the psychoanalysts have not bothered to work out its relations to the natural sciences.

One point which formerly troubled many psychologists was the treatment of goals or purposes. The idea that behavior

could be determined by purposes or goals which were achieved *after* the behavior was completed, seemed to indicate that the cause came after the effect. This, however, was an unnecessary assumption, since it is not eating the fish that leads a cat to escape from a problem box, but stomach contractions or some other anticipatory activity, e.g., vision, or even perhaps the imaginary smell of the fish, which leads to the act of escaping from the box.

With the progress of experimental science, more and more things which seemed to have no physical bases, have been shown to occur in strict accord with physical laws. Vitalism is a position no longer seriously held by any leading psychologists, and the question has now changed to that of refining the mechanistic statement so as to bring it into accord with present-day knowledge. The following definition gives an eclectic interpretation of the mechanistic view:

> Psychological phenomena are an integral part of the natural sciences, behaving in accordance with the orderly "laws" or principles of physics, chemistry, biology, etc., so far discovered and continuing to be extended along the same general lines. The doctrine of emergent evolution is consistent with this.

Question 4. *What are the respective roles of the analysis into fundamental components, and the analysis of the patterns in which psychological processes occur?*

The first kind of analysis seeks to trace all complex processes back to some simpler and irreducible basis similar to the physicist's analysis of matter into atoms, or later, electrons. Thus Wundt and Titchener sought to reduce all mental processes to a few simple types, the sensations and feelings, which were fundamental in that they could not be analyzed into anything simpler. They also happened to believe that these fundamental components were separate or discrete units, comparable to the elements of physics. However, these fundamental components might not be actually separate units, but might occur as continuous processes, or even as continuous gradations in kind or amount of some more general process,

e.g., vision. Although the existentialists accepted the view of separate units, and were roundly criticized by the functionalists and Gestaltists, none of them seem to have perceived the alternative viewpoint of continuous but equally fundamental processes. Likewise the behaviorists analyzed overt behavior into simpler reflexes, glandular and muscular, and again were criticized by the functionalists and Gestaltists for this view. In both cases the charge was that separate sensations or reflexes never were observed to occur in an intact, living organism, and that the simplest behavior ever observed was always a *pattern* of experience or behavior.

Perhaps the difficulty is that while physicists have recently invented methods to show that their fundamental components do occur in discrete units, psychology as yet has no comparable method to decide the question in this field. In this case we should suspend judgment.

What the functionalists and Gestaltists proposed as an alternative concept was that we should analyze experience or behavior only as far as it could be observed and verified with present methods. Thus, instead of analyzing a visual perception into separate sensations, which were only reported by observers long trained in these somewhat artificial procedures, we should accept the analyses given by typical careful, but untrained observers, who had not been prejudiced unintentionally in their expectations by Titchener's kind of training. Such analyses tended to indicate that experience is usually continuous, and that it is always patterned into a *figure* of greatest clearness, and a *ground* which is usually in the periphery and is consequently less clear. This type of analysis had already been done in a minor way by all psychologists, but the Gestaltists are bringing it to its highest development.

From the eclectic point of view, psychology should include *both* the analysis into fundamental components and the analysis of the characteristic patterns in which their components occur. Now that the old controversy about separate or continuous fundamental variables has cooled off somewhat, newer findings and methods may enable us to reopen this question of the

separateness or continuity of mental processes. By analogy a desk top appears to be continuous under a visual inspection, but is also believed to consist of separate electrons, etc., arranged in certain lace-like atomic and chemical patterns too minute to be observed with the unaided eye. Using examples such as this Gestalt psychologists have brought out the contrast between their emphasis upon the importance of the "field" or pattern and what they call the "constancy" or "bundle" hypothesis of the existentialists. The term constancy here refers to the idea that there are separate and unchanging mental elements which retain their individual characteristics, regardless of the patterns in which they appear. More refined methods of study might show that mental processes are not as continuous as they appear to be under ordinary observation.

Titchener classified perceptions into *mosaic* patterns in which details usually remain sharp, as in vision, and *fusions* or blends as in taste, smell, and hearing.

Organismic psychologists have often stated that "the whole is more than the sum of its parts." By this they mean that slight changes in the pattern of experience or behavior—e.g., including or omitting the word "please," or merely using a certain tone of voice in polite conversation—may almost completely reverse the significance of an action. This, however, causes no worry to other psychologists, who point out the analogy from chemistry, in which various slightly different *spatial arrangements* of *elements* are often sufficient to produce marked changes in the properties of a compound.

Psychoanalysis is an illustration of an analysis into a few *large* fundamental components, the ego, id, super ego, etc. It also emphasizes the importance of patterns of behavior, such as rationalization, phantasy, compensation, and a vast number of others.

Question 5. *What are the respective roles of static and dynamic analyses?*

A static analysis is a cross-sectional description of an experience or act at any one stage, treated as if it were unchanging at the moment. Conversely, a dynamic or longitudinal analysis

is one which traces a changing pattern of experience or action from the first to the last stages. This can be done only *by repeating or continuing the cross-sectional type of analysis.* A dynamic psychology is then merely one which emphasizes the importance of temporal changes in experience and action (which no one else would deny!). Such an interest is usually associated with a parallel emphasis upon pattern analysis, as mentioned in our example of the change in tone of voice in polite conversation. The same schools which emphasized analysis into fundamental components emphasized static analysis, and the ones which emphasized pattern analysis also emphasized longitudinal analyses. There may be a certain logical basis for the fact that the earlier points of view emphasized static analyses, because these were simpler and indicated the "lay of the land" as to the kinds of variables involved and a few typical patterns, whose development might then be traced by later investigators. This was then an example of division of labor, and after static analyses had been worked out by the first investigators, it was quite natural that later schools should take up where the earlier ones had left off. Again, eclectics would say that both types of analysis are essential to a complete psychology.

SUMMARY

In considering the standpoint of each school on each of these five systematic problems, we have seen (1) that the definition of the scope of psychology was largely a matter of specialized interests; (2) that various hypotheses as to the relation between mental and physical processes showed a progressive trend toward a unified description of both in a single scientific terminology; (3) that a mechanistic point of view can be restated to include the phenomena of purposes or goals, thus removing one of the last reasons for the adherence of any school to the outworn vitalistic position; (4a) that the so-called "atomistic" approach is one of analyzing the *fundamental components* of psychological processes, and as such cannot be discarded, and (4b) that the organismic approach simply analyzes

the *patterns* in which these fundamental components appear, differences between schools seeming to be largely differences in emphasis upon the importance of each type of analysis at various stages in the progress of psychology, both types of analysis being necessary to a complete scientific treatment; and (5) that much the same situation appears in the relative emphasis upon static or dynamic analyses. After the earlier schools had worked out the preliminary static analyses of any single temporal stage of a psychological activity, it was quite natural for later schools to emphasize the development of these same activities through the different stages from the earliest to the last.

Any attempt to answer all five questions in sufficient completeness to formulate a system would necessarily include the contributions of all viewpoints and fields of psychology. Instead of being logically opposed rival systems, these viewpoints may more adequately be considered as supplementary developments of a single eclectic system to which all have made significant contributions.

CHAPTER 41

Convergent Trends in Experimental Psychology[1]

In the previous chapter we have discussed five general problems of systematic psychology and the positions taken on each by representatives of five modern viewpoints or schools. These positions of each school represented their attempts to develop the general principles for a working program in the further development of psychology. Each such general principle was based primarily upon the directions of research they had found to be most significant in their own work, and which they consequently sought to extend over the whole field of psychology.

Now that we have shown the general working programs of each viewpoint, let us examine their contributions to the more specific problems of experimental psychology. Here again we must guard against confusion from the same four types of pseudo-controversy which we have previously mentioned: (1) arbitrary definition of the scope of psychology; (2) differences in emphasis on various aspects of problems; (3) use of different terminologies for the same aspects; and (4) evaluations of a system by "all-or-none" acceptance or rejection rather than on the specific merits and demerits of a given position. The division of labor by specialization on different aspects of a problem and the supplementary nature of the specialists' contributions should also be apparent here.

SENSORY PROBLEMS

Question 1. *What are the fundamental components of sensory experiences?*

[1] Authors cited in Ch. 40 will not be repeated in this chapter.

Titchener, after checking the experiments of Wundt and other existentialists, came to the conclusion that all complex sensory processes, e.g., perceiving a picture, imagining a concert, could be reduced to two fundamental sensory components, namely, sensations and images. Since these seemed to be irreducible components, he called them mental elements. Each element could be described with respect to four main characteristics or attributes, namely, (1) *quality,* e.g., redness in vision, sweetness in taste; (2) *intensity,* e.g., strength, loudness, weakness; (3) *protensity* (duration and intervals of time) and (4) *attensity* or clearness. In addition, some sense fields, such as vision, touch, and kinesthesis, may have a fifth characteristic or attribute of *extensity,* or spatial arrangement.

Titchener found that visual images differed from visual perceptions primarily in their characteristic lower degree of clearness. He personally believed this characteristic difference to be great enough to distinguish images as a separate type of element, just as some chemical elements which resemble one another very closely differ in some minor but definite characteristic. He also recognized that others might legitimately consider images to be essentially the same as sensations, as they have since done. One of his students, Perky, found that observers who were in a dark room and looking at a neutral gray screen would report that their imagery was unusually distinct on certain days, when unknown to them, a very dim lantern slide of the same object was projected on the screen from the back. Perky also found that when the observers were later asked to describe their perceptions of these same but slightly brighter lantern slide figures they would often go on describing their perceptions after the projector had been turned off for many seconds. Evidently the sensation and image derived from the original source are so alike that observers have great difficulty in distinguishing them.

Others who have repeated this kind of experiment have in general confirmed the fact that sensory phenomena can be analyzed in terms of these five attributes, except that clearness is usually regarded as being analyzable into *relative* differences

in intensity. The few controversies regarding the interpretation of such experiments have to do with whether other attributes such as "tonality" are secondary complex attributes or whether they too are psychologically fundamental. Most investigators have considered all other attributes to be secondary.

Behaviorists who originally tried to scrap all mentalistic terms have subsequently discovered that the term "discriminative response" from Pavlov's work on animals [2] is a general equivalent to all sensory processes. Interestingly enough, Boring, Langfeld and Weld,[3] previously close associates of Titchener, have recently stated in their introductory text that the term discriminative response seems to be the basic descriptive unit for psychology. Aside from the previously mentioned debate on the constancy hypothesis, and other minor differences in interpreting experiments, we now have fairly close agreement on this problem.

Question 2. *What are the characteristic patterns in which sensory processes occur?*

Titchener distinguished three characteristic patterns of complex sensory experiences as follows: (1) perceptions, those in which sensations (brought about by present stimulation) were focal, with images and feelings in the margin of clearness; (2) memory images, or reproductions of previous perceptions, and (3) imagination, or recombination of imagery patterns from previous experiences in a way which did not conform to any actual previous situation.

It had long been known and experimentally demonstrated that the setting for an experience might have a marked effect upon our perceptions of any part of a given situation. A whole psychological movement called "apperceptionism" arose and later fell from over-elaboration on this point. The Gestaltists, however, raised the question as to whether sensory patterns were built up from separate sensations or whether new-born

[2] Pavlov, I. P., *Conditioned Reflexes:* An Investigation of the Physiological Activity of the Cortex, London, Oxford University Press, 1927, pp. 430.

[3] Boring, E. G., Langfeld, H. S., and Weld, H. P., *Psychology:* A Factual Textbook, New York, Wiley, 1935, pp. 555.

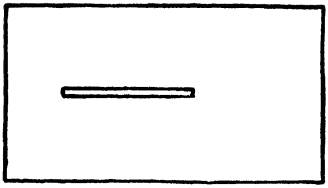
organisms did not respond in a unified way from the very beginning of activity and only develop part responses such as reflexes or sensations as a result of later *differentiation.*

An example of the way in which a total situation may determine a perception without the observer's having any previous knowledge of its separate parts is given in a Gestalt experiment on apparent movement (the *phi* phenomenon). Fig. 70, Card B, represents a rectangular cardboard having two slits, the entire front being covered by a sheet of tissue paper to obscure any details under ordinary light. If now Card A having a single horizontal slot is moved up and down back of the two slits, at various times a light from in back of both cards will shine through the openings and show up as one or more lighted bars.

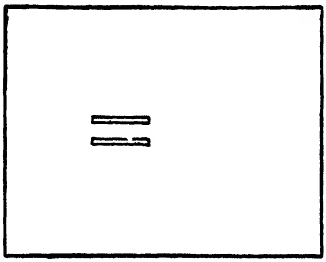
Experimentally it is found that if the back card is moved very slowly, we see the lower and upper bars separately in alternation. As the rate of fluctuation is increased to a certain point we see a single bar moving across the face of the card from the lower to the upper positions at a regular rate. Since the lights do not actually move, this is called apparent movement. This same principle of stroboscopic vision is the basis for ordinary moving pictures.

Does an observer then perceive each bar separately and unconsciously "infer" that this represents a single bar moving between the two positions? Observers who see it for the first time at the optimum rate without knowledge of the apparatus do not report any such process of inference. They simply see the moving bar. If there is an unconscious inference it should be possible to find the cues for such interpretation. Korte, a Gestalt psychologist, did determine the optimum size, light intensity, duration, and interval between light flashes as they affected apparent movement. His experiments have been extended by De Silva.[4] The previously mentioned example of

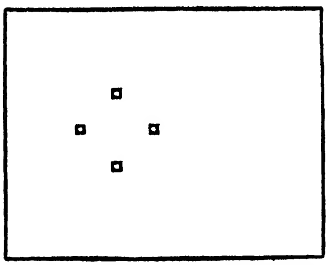
[4] De Silva, H. R., "An Analysis of the Visual Perception of Movement," *Brit. J. of Psychol., 19,* pt. 3, 1929, pp. 269-273, 275, 291. Also as Reading 21 in Wheeler, R. H., *Readings in Psychology,* New York, Crowell, 1930, pp. 411-433.



Card A—6 × 11 in. Slit ¼ × 4½ in. equidistant from top and bottom, and 2 in. from left-hand edge.



Card B—8½ × 11 in. Slits $\frac{3}{16}$ × 2 in. spaced ½ in. apart and 3 in. from left-hand edge.



Card C—8½ × 11 in. Holes ¼ in. square, spaced 1½ in. apart and the nearest one 3 in. from left-hand edge.

FIG. 70. Material for demonstration of the phi-phenomenon. (From Seashore and Seashore, *Elementary Experiments in Psychology*, Holt, 1935, p. 114.)

Korte's laws, on the optimal conditions for apparent movement, represent an initial attack upon the supplementary problem of analyzing pattern phenomena.

Behaviorists would simply say that in the perception of apparent movement we have cues similar to those in ordinary perception of actual movement, so that through conditioning or redintegration [5] a few of these cues are alone sufficient to produce the impression of movement without the observer's having to have all of the cues as in actual movement. Therefore we would not necessarily be aware of any such short-cut type of inference, any more than we are aware of the separate kinesthetic cues necessary to maintain our balance after learning to stand erect in early childhood.

One incidental explanation, that alternate flashes would cause incipient eye movements, was studied by having four squares arranged in a diamond pattern as in Card C. In these experiments observers could see a square move from position 1 to position 3 to position 2, through either a right-angled path or a curved path. Furthermore, one seemed to see the square apparently split and move *simultaneously in opposite directions* to both positions 3 and 4, and then converge on position 2, reverse the process and continue on again. Observers could even see both right and left directions of apparent movement at the same time, *with only one eye open!* Could one eye move in opposite directions at the same time, and could the apparent movement rest upon eye movement cues as suggested by the behaviorists and others? Unfortunately this demonstration does not eliminate the importance of eye movements, because it would be possible to have opposing eye muscles contract at the same time and produce kinesthetic cues from *strain* rather than movement. Perhaps electrophysiological recordings of ac-

[5] Sir William Hamilton, a Scottish psychologist, long ago pointed out that whereas a whole pattern of stimuli, e.g., a printed song, might be necessary to set off the action (of singing), later on through experience or "redintegration" a fragment of the original pattern might be sufficient to produce the response. This long-neglected principle has also been further developed in more modern texts by Hollingworth, and is described in the work of Murphy cited in the previous chapter.

tion currents from separate eye muscles may give the answer.

We may agree that changes in the spatial and temporal patterning of stimuli can produce just as important effects as changing any of the parts alone. The Gestaltists have done a great deal to emphasize the importance of the patterning of psychological processes, both spatially and temporally. An eclectic might point out that no one doubts we can learn to differentiate our early gross reactions into finer part reactions, as in the infant's learning to grasp and manipulate objects with its fingers. This in no way contradicts the earlier principle that we can also combine experiences or overt acts to secure more complex reactions. Both differentiation from larger to finer reactions and integration from simpler to more complex reactions are found to occur in everyday life.

Question 3. *What are some of the typical functional principles in sensory activities?*

Here we need only cite some of the laws of complementary colors as applied to color mixture, contrast, and after images to demonstrate the permanent value of the early structuralist contribution to psychology.

AFFECTIVE PROCESSES

Question 1. *What are the fundamental (irreducible) components of affective processes (feelings, emotions, etc.)?*

Here again the structuralists who first analyzed the content of such experience have given us our principal information. After many introspective reports by trained observers, Titchener finally concluded that there is only one type of elementary process, which he called *simple feelings*. For these he listed the same four attributes as for sensation and image, namely, quality, intensity, protensity or time, and attensity or clearness. Feelings, he held, have only two opposed qualities, pleasantness and unpleasantness, and characteristically tend to become less clear when closely examined. It may eventually turn out that feelings are simply organic sensations from the viscera; Carr's functional theory accepts many of Titchener's findings, but holds that pleasantness may be a by-product of a positive

response to a situation and unpleasantness an aspect of negative or withdrawal responses. However, he recognizes that the explanation is not that simple and that the situation may be reversed by conditioning.

Watson had no interest in any such explanations and considered emotions to be simply unorganized visceral activity of muscles and glands. Psychoanalysts have had little to say about elementary feelings except to classify them into a broad group, the life instinct, and an opposite, usually unpleasant group, the death instinct. This appears to be purely a clinical device for explaining a patient's difficulties to him, and is one of the lesser aspects of their system.

An eclectic might point out that Titchener seems to be essentially correct as to the kinds of simplest feelings, that Carr's theory is not incompatible with it, that both may be aspects of Watson's theory of visceral activity, and that the psychoanalysts' theory is not yet clear enough to be evaluated as a part of an experimental body of facts.

Question 2. *What are the characteristic patterns of affective processes?*

Here we may mention a long list of overlapping concepts such as emotion, interest, attitude, motivation, conflict, and all of the many Freudian mental mechanisms such as rationalization and identification.

Titchener defined emotions as experiences which had a focal core of simple feelings on a mixed background of sensations and images, and which followed a characteristic temporal course in their rise and fall. Such topics as interests, attitudes, and motivation have been studied only rather recently and usually by eclectics or applied psychologists for some immediate purpose, rather than for systematic classification. Here we may recall the work of E. K. Strong, Jr., on interests (Ch. 25), and Thurstone on measuring social attitudes (Ch. 35).

The Gestaltists have done a great deal to demonstrate the importance of goals as motivation in learning, and a related group of topologists, led by Kurt Lewin, has done a great deal to devise experiments for the analysis of behavior in conflict

situations. Topologists such as Lewin, and behaviorists such as Hull and his colleagues at Yale are at last finding ways to investigate experimentally the nature of the adjustment mechanisms discovered in Freud's clinical work on mental disorders.[6] These experiments typically begin by placing some person or group in a conflict situation where their ordinary responses result only in frustration. Usually the persons then resort to the Freudian type of adjustment mechanisms, such as rationalization, compensation, or aggression.

Question 3. *What are some typical functional principles in the field of affective processes?* Here we might list only a few from each group.

a. (Existentialists) Simple feelings of pleasantness and unpleasantness may follow one another very closely, but they are never present together.

b. (Functionalists) Pleasantness is related to positive, unpleasantness to negative responses.

c. (Behaviorists, Sherman and Sherman's revision of Watson's work) Emotions of human infants are relatively undifferentiated at birth, but become somewhat differentiated through experience.

d. (Gestaltists) Learning is always goal directed.

e. (Psychoanalysts) All behavior is motivated by and often distorted by desires of which we are not clearly aware.

MOTOR PROCESSES (OVERT BEHAVIOR)

Question 1. *What are the fundamental components of motor activity?*

Here we find that the existentialists simply did not consider behavior study to be a part of psychology. They were interested in it only as an opportunity to study "the action consciousness," or kinesthesis and equilibrium. John Dewey wrote one of the earliest functionalist articles on a warning against the dangers of considering reflexes to be separate units, which he regarded as a parallel to the danger of regarding sensations or images as mental elements. He proposed instead the study

[6] Cf. the experiments of Lewin and Hovland and Sears (Ch. 34).

of complete acts of adjustment as the simplest descriptive unit.

The behaviorists have followed physiologists in considering reflexes (or tropisms in the case of some of the lower animals) to be the fundamental components of muscular and glandular aciton.

Gestaltists, like the functionalists, have simply given up this problem, and directed their attention to the second problem, patterning of action. Psychoanalysts have not been interested in the problem at all.

Question 2. *What are the characteristic patterns of overt behavior?* In this relatively undeveloped field we find no particular differences between those schools which are interested and only rough attempts at classification by any of them, e.g., gross bodily co-ordinations such as athletic skills vs. fine muscle co-ordinations such as the manual and finger skills required in typewriting.

Both fine and gross skills may be further analyzed as to: (a) quality (kind of movements, e.g., manual, postural, walking, speaking, etc.); (b) intensity (strength); (c) time (duration, rate, intervals); and (d) extensity (spatial precision). Experts in the industrial engineering field of time and motion study have discovered a list of eighteen relatively simple actions which go to make up most of our ordinary manual skills, such as factory operations. These units or "therbligs" (after their discoverer, Gilbreth) are more complex than reflexes, but are fairly distinct and easily identified by ordinary observation or through moving pictures.

Question 3. *What are some typical functional principles concerning motor processes?* Two answers stand out:

a. The functionalist and behaviorist theory that all sensory activity leads to some type and degree of motor activity affords a possible answer to the mind-body problem in stating that all mental processes are forms of implicit behavior.

(b) The motion study principles of utilizing simultaneous symmetrical movements of both hands whenever possible.

THOUGHT PROCESSES

Question 1. *What are the fundamental components of thought?*

In a prolonged series of experiments and in the theoretical controversy between the Titchener (Cornell) and the Külpe (Würzburg) groups, Titchener and his highly trained observers maintained that all thinking could be introspectively analyzed into series of images, and that there was no separate thought element. Külpe, using simpler introspective reports from untrained observers, found many instances in which no images were reported, and others of his students believed that in the absence of such images there must be a separate thought element. Both groups agreed that the preliminary mental set or determining tendency was very important in bringing about the solution to a thought problem.

Perhaps this controversy may be partly cleared up by the following considerations: (a) People differ widely in the clearness of their imagery in various fields, and many people have never recognized the possibility of imagery in other than visual or auditory fields. Hence some naive introspectors may have failed to analyze their thinking far enough to detect relatively faint imagery, especially kinesthetic imagery from speech muscles, which Titchener's trained observers found to be very important. (b) In problem solving, involving a series of steps, the process may be slow enough to enable observers to recognize the importance of imagery in various imaginary trial and error steps. But if a problem of this type is practiced very often, we develop short cuts from the recognition of small or faint cues and respond almost automatically. In this case it might be quite difficult to observe and describe any conscious process.

Thus thinking of some sorts may proceed via the use of imagery, while in other simpler and more automatic processes imagery is not so apparent, especially in untrained observers and people who have less vivid imagery in the dominant sense fields, such as vision.

The behaviorists would consider thinking as implicit speech activity, and hence as a motor process, a form of conditioned reflexes. Gestaltists reject the idea of any thought unit smaller than a pattern, while the psychoanalysts again have not been interested in this problem.

Question 2. *What are the characteristic patterns of thinking?*

Here again the study of thinking, although one of the oldest in psychology, is still relatively unexplored by the experimental method. As common pattern names, we may find the overlapping terms of problem solving, judgment, thinking, and reasoning, both inductive and deductive.

Question 3. *What are some typical functional principles of thinking?*

Existentialists have described thinking as a series of associations of mental elements, with elaborate classifications of primary and secondary laws of association, such as contiguity, recency and frequency. Functionalists have studied thinking as a means of adjustment to environment, while Behaviorists have attempted to show that it is simply language behavior, usually implicit vocal movement in a person who can speak, but gesture or the finger activity of sign language in a deaf mute individual. Psychoanalysts have shown many examples of the way in which unconscious or at least not clearly conscious motivation distorts and directs our thinking. From their viewpoint, thinking is not the dominant aspect of our life, but a very corruptible servant dominated by our motives.

ADDITIONAL TOPICS

Any well-rounded picture of various viewpoints in experimental psychology would need to include at least three other topics of current interest. Fortunately our previous sections have already touched upon these and so we need only state the questions and a few general trends.

Topic 1. *What is the organization of human abilities, as to general, group, and specific factors?*

The chapters on factor analysis (5, 20) have outlined

progress in this area. In general, existentialists and behaviorists have emphasized the role of specific factors, functionalism and psychoanalysis the role of group factors, and Gestalt psychology the general or unitary functioning of the organism. Further progress in this field seems to depend in part upon the acquisition and analysis of intercorrelations between various human characteristics. This type of work now represents a fairly large percentage of psychological research. The other biggest need is the careful redefinition of terms and concepts to clear up misunderstandings.

Topic 2. *Is there a simple order of development in behavior with regard to differentiation and integration?*

Here the studies in comparative psychology have indicated that the earliest stages of behavior do not progress in the same way for all organisms at all times. In some fetal organisms, relatively specific reflexes may be found before any more widespread action is observed, while in other organisms more general activity precedes the development of specific movements of separate reflexes. Apparently, then, we may accept both integration of reflexes to more complex movements and differentiation of general into specific movements, and there is no apparent reason why both integration and differentiation may not be important at different stages of the development of each organism. We can either combine or refine our actions.

Such an eclectic summary would resolve one of the disputes between Gestalt, or organismic psychology, and behaviorism.

Topic 3. *What are the essential factors in learning with regard to (a) the nature of the activity, random or purposive, and (b) the rate of learning, sudden insight or trial and error?*

The main points involved in the answers to these questions are given in a recent publication by the writer.[7] The conclusion here presented, and in keeping with a similar analysis by Dashiell,[8] is that both mechanistic and organismic interpreta-

[7] Seashore, R. H., The Pyramid Puzzle: A Useful Device in Studying Thought. *Amer. J. Psychol.*, 1938, *51*, 549-557.

[8] Dashiell, J. F., *Fundamentals of General Psychology*, New York, Houghton Mifflin, 1937.

tions are part of a larger and more general eclectic theory of learning to which both viewpoints have contributed very significantly. Such a statement might well be made of the contributions of various schools of thought to almost every other major problem in psychology.

Index